GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:38:14 ; Search time 22.23 Seconds
(without alignments)
77.805 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 11821

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	34.6	23	2	A59048	convulsant peptide
2	36	34.6	46	1	DKDCB	denclatoxin B - Co
3	34	32.7	23	2	E39855	paralytic peptide
4	34	32.7	23	2	C39855	paralytic peptide
5	34	32.7	23	2	D39855	paralytic peptide
6	33.5	32.2	25	2	A58647	alphaA-conotoxin P
7	33.5	32.2	48	2	S29216	neurotoxin Tx2 - s
8	32	30.8	23	2	F39855	paralytic peptide
9	32	30.8	23	2	G39855	paralytic peptide
10	31.5	30.3	42	2	A31918	cathepsin D (EC 3.
11	31	29.8	50	2	D72804	gp38 protein - Myc
12	30	28.8	47	2	G81008	hypothetical prote
13	30	28.8	50	2	H90760	hypothetical prote
14	29	27.9	19	2	S62864	toxin VI - Tityus
15	29	27.9	23	2	I53401	monocyte chemotact
16	29	27.9	29	2	A56283	kalata B1 [validat
17	29	27.9	34	2	I65263	homeobox protein H
18	29	27.9	36	2	S75704	plantaricin C19 -
19	28.5	27.4	26	2	S55029	CAP3 protein - ant
20	28.5	27.4	44	2	S29975	hypA protein - Alc
21	28	26.9	14	2	I56493	endothelial growth
22	28	26.9	27	4	S53259	probable pre-core
23	28	26.9	41	2	S19566	ornatin A2 - leech
24	28	26.9	44	2	I48942	cellular disintegr
25	28	26.9	45	2	F90716	probable RNA [impo
26	28	26.9	45	2	F64801	hypothetical prote
27	28	26.9	48	2	D90777	hypothetical prote
28	28	26.9	48	2	S42399	hypothetical prote
29	28	26.9	48	2	E85646	hypothetical prote
30	27	26.0	23	2	I54773	neural cell adhesi
31	27	26.0	23	2	A60226	pyruvate dehydroge
32	27	26.0	25	2	JH0700	omega-conotoxin MV
33	27	26.0	33	2	A36154	benzphetamine N-de
34	27	26.0	37	1	A42040	kaliotoxin 1 [vali
35	27	26.0	38	2	A54471	agitoxin 1 - scorp
36	27	26.0	38	2	B54471	agitoxin 2 - scorp
37	27	26.0	38	2	C54471	agitoxin 3 - scorp
38	27	26.0	39	2	C97513	hypothetical prote
39	27	26.0	40	2	JT0515	Ig heavy chain V-I
40	27	26.0	41	2	T46621	hypothetical prote
41	27	26.0	43	2	B41711	defensin B - beetl
42	27	26.0	44	2	S05017	alpha-amylase inhi
43	27	26.0	46	2	I48947	cellular disintegr
44	27	26.0	46	2	A83629	hypothetical prote
45	27	26.0	46	2	C83437	hypothetical prote

ALIGNMENTS

RESULT 1
A59048
convulsant peptide - cone shell (Conus textile)
C;Species: Conus textile (cloth-of-gold cone)
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
C;Accession: A59048
R;Cruz, L.J.; Ramilo, C.A.; Corpuz, G.P.; Olivera, B.M.
Biol. Bull. 183, 159-164, 1992
A;Title: Conus peptides: phylogenetic range of biological activity.
A;Reference number: A59048
A;Accession: A59048
A;Molecule type: protein
A;Residues: 1-23 <CRU>
C;Keywords: amidated carboxyl end; neurotoxin; venom
F;23/Modified site: amidated carboxyl end (Pro) #status predicted

Query Match 34.6%; Score 36; DB 2; Length 23;
Best Local Similarity 62.5%; Pred. No. 46;
Matches 5; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy       9 CSXXGCRP 16
           |  ||||
Db      15 CEASGCRP 22
```

RESULT 2
DKDCB
denclatoxin B - Columbia mistletoe
C;Species: Dendrophthora clavata (Columbia mistletoe)
C;Date: 30-Apr-1981 #sequence_revision 30-Apr-1981 #text_change 04-Oct-1996
C;Accession: A01804
R;Samuelsson, G.; Pettersson, B.
Acta Pharm. Suec. 14, 245-254, 1977
A;Title: Toxic proteins from the mistletoe Dendrophthora clavata.
A;Reference number: A01804; MUID:78016835
A;Accession: A01804
A;Molecule type: protein
A;Residues: 1-46 <SAM>
C;Superfamily: viscotoxin
C;Keywords: toxin
F;3-40,4-32,16-26/Disulfide bonds: #status predicted

Query Match 34.6%; Score 36; DB 1; Length 46;
Best Local Similarity 66.7%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      10 SXXGCRPGY 18
           |  || |||
Db      36 SGTGCPPGY 44
```

```
RESULT   3
E39855
paralytic peptide III - beet armyworm
C;Species: Spodoptera exigua (beet armyworm)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 30-Sep-1993
C;Accession: E39855
R;Skinner, W.S.; Dennis, P.A.; Li, J.P.; Summerfelt, R.M.; Carney, R.L.; Quistad, G.B.
J. Biol. Chem. 266, 12873-12877, 1991
A;Title: Isolation and identification of paralytic peptides from hemolymph of the lepidc
A;Reference number: A39855; MUID:91302298
A;Accession: E39855
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-23 <SKI>
C;Superfamily: paralytic peptide I

  Query Match             32.7%;  Score 34;  DB 2;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 91;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   4
C39855
paralytic peptide I - beet armyworm
C;Species: Spodoptera exigua (beet armyworm)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 30-Sep-1993
C;Accession: C39855
R;Skinner, W.S.; Dennis, P.A.; Li, J.P.; Summerfelt, R.M.; Carney, R.L.; Quistad, G.B.
J. Biol. Chem. 266, 12873-12877, 1991
A;Title: Isolation and identification of paralytic peptides from hemolymph of the lepidc
A;Reference number: A39855; MUID:91302298
A;Accession: C39855
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-23 <SKI>
C;Superfamily: paralytic peptide I

  Query Match             32.7%;  Score 34;  DB 2;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 91;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   5
D39855
paralytic peptide II - beet armyworm
C;Species: Spodoptera exigua (beet armyworm)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 30-Sep-1993
C;Accession: D39855
R;Skinner, W.S.; Dennis, P.A.; Li, J.P.; Summerfelt, R.M.; Carney, R.L.; Quistad, G.B.
J. Biol. Chem. 266, 12873-12877, 1991
A;Title: Isolation and identification of paralytic peptides from hemolymph of the lepidc
A;Reference number: A39855; MUID:91302298
A;Accession: D39855
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-23 <SKI>
C;Superfamily: paralytic peptide I

  Query Match             32.7%;  Score 34;  DB 2;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 91;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   6
A58647
alphaA-conotoxin PIVA [validated] - cone shell (Conus purpurascens)
C;Species: Conus purpurascens (purple cone)
C;Date: 31-Oct-1997 #sequence_revision 07-Nov-1997 #text_change 15-Sep-2000
C;Accession: A58647
R;Hopkins, C.; Grilley, M.; Miller, C.; Shon, K.J.; Cruz, L.J.; Gray, W.R.; Dykert, J
J. Biol. Chem. 270, 22361-22367, 1995
A;Title: A new family of Conus peptides targeted to the nicotinic acetylcholine rece
A;Reference number: A58647; MUID:95403432
A;Accession: A58647
A;Molecule type: protein
A;Residues: 1-25 <HOP>
R;Han, K.H.; Hwang, K.J.; Kim, S.M.; Kim, S.K.; Gray, W.R.; Olivera, B.M.; Rivier, J
submitted to the Brookhaven Protein Data Bank, December 1996
A;Reference number: A67666; PDB:1P1P
A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR, res
R;Han, K.H.; Hwang, K.J.; Kim, S.M.; Kim, S.K.; Gray, W.R.; Olivera, B.M.; Rivier, J
Biochemistry 36, 1669-1677, 1997
A;Title: NMR structure determination of a novel conotoxin, [Pro 7,13] alpha A-conoto
A;Reference number: A58646; MUID:97200721
A;Contents: annotation; conformation and disulfide bond assignments by (1)H-NMR
C;Superfamily: unassigned conotoxins
C;Keywords: acetylcholine receptor inhibitor; amidated carboxyl end; hydroxyproline;
F;2-16,3-11,14-23/Disulfide bonds: #status experimental
F;7,13/Modified site: 4-hydroxyproline (Pro) (partial) #status experimental
F;20/Modified site: 4-hydroxyproline (Pro) #status experimental
F;25/Modified site: amidated carboxyl end (Gln) #status experimental

  Query Match             32.2%;  Score 33.5;  DB 2;  Length 25;
  Best Local Similarity   57.1%;  Pred. No. 1.2e+02;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    5;  Gaps    2;

Qy       7 HPCSXXGC--RPGY 18
           ||||   |  || |
Db      12 HPCS---CKDRPSY 22


RESULT   7
S29216
neurotoxin Tx2 - spider (Phoneutria nigriventer)
C;Species: Phoneutria nigriventer
C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 07-May-1999
C;Accession: S29216
R;do Nascimento Cordeiro, M.; Ribeiro Diniz, C.; do Carmo Valentim, A.; von Eicksted
FEBS Lett. 310, 153-156, 1992
A;Title: The purification and amino acid sequences of four Tx2 neurotoxins from the
A;Reference number: S29214; MUID:93011905
A;Accession: S29216
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-48 <COR>
C;Superfamily: curtatoxin

  Query Match             32.2%;  Score 33.5;  DB 2;  Length 48;
  Best Local Similarity   41.2%;  Pred. No. 2.1e+02;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    3;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPGY 18
           : : | ||      || ||
Db      22 ECVCGGPCI---CRQGY 35


RESULT   8
F39855
```

paralytic peptide I - tobacco budworm
C;Species: Heliothis virescens (tobacco budworm)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 30-Sep-1993
C;Accession: F39855
R;Skinner, W.S.; Dennis, P.A.; Li, J.P.; Summerfelt, R.M.; Carney, R.L.; Quistad, G.B.
J. Biol. Chem. 266, 12873-12877, 1991
A;Title: Isolation and identification of paralytic peptides from hemolymph of the lepidc
A;Reference number: A39855; MUID:91302298
A;Accession: F39855
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-23 <SKI>
C;Superfamily: paralytic peptide I

Query Match 30.8%; Score 32; DB 2; Length 23;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 GCRPGY 18
|| |||
Db 6 GCIPGY 11

RESULT 9
G39855
paralytic peptide II - tobacco budworm
C;Species: Heliothis virescens (tobacco budworm)
C;Date: 30-Dec-1991 #sequence_revision 30-Dec-1991 #text_change 30-Sep-1993
C;Accession: G39855
R;Skinner, W.S.; Dennis, P.A.; Li, J.P.; Summerfelt, R.M.; Carney, R.L.; Quistad, G.B.
J. Biol. Chem. 266, 12873-12877, 1991
A;Title: Isolation and identification of paralytic peptides from hemolymph of the lepidc
A;Reference number: A39855; MUID:91302298
A;Accession: G39855
A;Status: preliminary
A;Molecule type: protein
A;Residues: 1-23 <SKI>
C;Superfamily: paralytic peptide I

Query Match 30.8%; Score 32; DB 2; Length 23;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 13 GCRPGY 18
|| |||
Db 6 GCIPGY 11

RESULT 10
A31918
cathepsin D (EC 3.4.23.5) - bovine (fragment)
C;Species: Bos primigenius taurus (cattle)
C;Date: 21-May-1990 #sequence_revision 31-Dec-1991 #text_change 01-Nov-1996
C;Accession: A31918
R;Yonezawa, S.; Takahashi, T.; Wang, X.; Wong, R.N.S.; Hartsuck, J.A.; Tang, J.
J. Biol. Chem. 263, 16504-16511, 1988
A;Title: Structures at the proteolytic processing region of cathepsin D.
A;Reference number: A92681; MUID:89034127
A;Accession: A31918
A;Molecule type: protein
A;Residues: 1-42 <YON>
C;Superfamily: pepsin
C;Keywords: aspartic proteinase; glycoprotein; hydrolase; lysosome
F;1-30/Product: cathepsin D light chain (fragment) #status experimental <LCH>
F;18-42/Product: cathepsin D, single-chain form (fragment) #status experimental <MAT>
F;33-42/Product: cathepsin D heavy chain (fragment) #status experimental <HCH>
F;1,28/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 30.3%; Score 31.5; DB 2; Length 42;
Best Local Similarity 41.2%; Pred. No. 3.7e+02;
Matches 7; Conservative 2; Mismatches 7; Indels 1; Gaps 1;

Qy 1 QDTIHGHPCSXXGCRPG 17
|||: ||: ||
Db 20 QDTV-SVPCNPSSSSPG 35

RESULT 11
D72804
gp38 protein - Mycobacterium phage D29
C;Species: Mycobacterium phage D29
C;Date: 12-Nov-1999 #sequence_revision 12-Nov-1999 #text_change 20-Apr-2001
C;Accession: D72804
R;Ford, M.E.; Sarkis, G.J.; Belanger, A.E.; Hendrix, R.W.; Hatfull, G.F.
J. Mol. Biol. 279, 143-164, 1998
A;Title: Genome structure of mycobacteriophage D29: Implications for phage evolution.
A;Reference number: A72800; MUID:98300335
A;Accession: D72804
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <FOR>
A;Cross-references: GB:AF022214; NID:g3172250; PIDN:AAC18479.1; PID:g3172286
C;Genetics:
A;Gene: 38

Query Match 29.8%; Score 31; DB 2; Length 50;
Best Local Similarity 50.0%; Pred. No. 5.1e+02;
Matches 5; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 9 CSXXGCRPGY 18
| | |||
Db 20 CDGGGSAPGY 29

RESULT 12
G81008
hypothetical protein NMB2072 [imported] - Neisseria meningitidis (strain MC58 serogro
C;Species: Neisseria meningitidis
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C;Accession: G81008
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Masignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A;Reference number: A81000; MUID:20175755
A;Accession: G81008
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-47 <TET>
A;Cross-references: GB:AE002557; GB:AE002098; NID:g7227332; PIDN:AAF42391.1; PID:g722
A;Experimental source: serogroup B, strain MC58
C;Genetics:
A;Gene: NMB2072

Query Match 28.8%; Score 30; DB 2; Length 47;
Best Local Similarity 50.0%; Pred. No. 6.8e+02;
Matches 5; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 6 GHPCSXXGCR 15
| || || |
Db 6 GKPCRPSRCR 15

RESULT 13
H90760
hypothetical protein ECs1056 [imported] - Escherichia coli (strain O157:H7, substrain
C;Species: Escherichia coli
C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C;Accession: H90760

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A;Reference number: A99629; MUID:21156231; PMID:11258796
A;Accession: H90760
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-50 <HAY>
A;Cross-references: GB:BA000007; PIDN:BAB34479.1; PID:g13360516; GSPDB:GN00154
A;Experimental source: strain O157:H7, substrain RIMD 0509952
C;Genetics:
A;Gene: ECs1056

Query Match 28.8%; Score 30; DB 2; Length 50;
Best Local Similarity 50.0%; Pred. No. 7.2e+02;
Matches 5; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

```
Qy       5 HGHPCSXXGC 14
           :| |||   |
Db      10 NGMPCSLNWC 19
```

RESULT 14
S62864
toxin VI - Tityus bahiensis (fragment)
C;Species: Tityus bahiensis
C;Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 07-May-1999
C;Accession: S62864
R;Becerril, B.; Corona, M.; Coronas, F.I.V.; Zamudio, F.; Calderon-Aranda, E.S.; Fletche
Biochem. J. 313, 753-760, 1996
A;Title: Toxic peptides and genes encoding toxin gamma of the Brazilian scorpions Tityus
A;Reference number: S62861; MUID:96190713
A;Accession: S62864
A;Molecule type: protein
A;Residues: 1-19 <BEC>
C;Superfamily: scorpion neurotoxin
C;Keywords: neurotoxin; venom

Query Match 27.9%; Score 29; DB 2; Length 19;
Best Local Similarity 40.0%; Pred. No. 4.3e+02;
Matches 4; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

```
Qy       6 GHPCSXXGCR 15
           |:|    ||:
Db       4 GYPTDKRGCK 13
```

RESULT 15
I53401
monocyte chemotactic protein - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 21-Jul-2000
C;Accession: I53401
R;Steenbergen, E.J.; Verhagen, O.J.; van Leeuwen, E.F.; Behrendt, H.; Merle, P.A.; Weste
Eur. J. Immunol. 24, 900-908, 1994
A;Title: B precursor acute lymphoblastic leukemia third complementarity-determining regi
fetal life.
A;Reference number: I53401; MUID:94200227
A;Accession: I53401
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-23 <RES>
A;Cross-references: GB:S69742; NID:g546303; PIDN:AAD14040.1; PID:g4261740
C;Genetics:
A;Gene: IgH-VDJ

Query Match 27.9%; Score 29; DB 2; Length 23;
Best Local Similarity 46.2%; Pred. No. 5.1e+02;
Matches 6; Conservative 0; Mismatches 5; Indels 2; Gaps 1;

```
Qy       6 GHP--CSXXGCRP 16
           | |  ||   | |
Db       3 GQPPYCSSTSCYP 15
```

Search completed: August 26, 2002, 13:38:15
Job time: 239 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:43:29 ; Search time 17.57 Seconds
(without alignments)
39.667 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 3667

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	34.6	46	1	THN_DENCL	P01541 dendrophtho
2	34	32.7	23	1	CP23_SPOER	P56683 spodoptera
3	34	32.7	23	1	PAP1_SPOEX	P30255 spodoptera
4	34	32.7	23	1	PAP2_SPOEX	P30256 spodoptera
5	34	32.7	23	1	PAP3_SPOEX	P30257 spodoptera
6	33.5	32.2	25	1	CXA4_CONPU	P55963 conus purpu
7	32	30.8	23	1	PAP1_HELVI	P30251 heliothis v
8	32	30.8	23	1	PAP2_HELVI	P30252 heliothis v
9	31	29.8	21	1	DCMS_PSECA	P19921 pseudomonas
10	31	29.8	30	1	VARB_VIOAR	P58447 viola arven
11	31	29.8	30	1	VARG_VIOAR	P58452 viola arven
12	31	29.8	30	1	VARH_VIOAR	P58453 viola arven
13	31	29.8	50	1	VG38_BPMD2	O64229 mycobacteri
14	29	27.9	19	1	SCX6_TITBA	P56610 tityus bahi
15	29	27.9	29	1	CYOC_VIOOD	P58444 viola odora
16	29	27.9	29	1	KABS_OLDAF	P58458 oldenlandia
17	29	27.9	29	1	VARA_VIOAR	P58446 viola arven
18	29	27.9	29	1	VARC_VIOAR	P58448 viola arven
19	29	27.9	29	1	VARD_VIOAR	P58449 viola arven
20	29	27.9	29	1	VARE_VIOAR	P58450 viola arven
21	29	27.9	37	1	TX21_SELHU	P82959 selenocosmi
22	29	27.9	37	1	TX22_SELHU	P82960 selenocosmi
23	28.5	27.4	26	1	MT1_COLGL	Q99334 colletotric
24	28.5	27.4	44	1	HYPA_ALCEU	P31901 alcaligenes
25	28	26.9	41	1	ORN2_PLAOR	P25509 placobdella
26	28	26.9	48	1	YO48_BPHK0	Q37928 bacteriopha
27	27	26.0	25	1	CXOA_CONMA	P05484 conus magus
28	27	26.0	38	1	SCA1_LEIQH	P46110 leiurus qui
29	27	26.0	38	1	SCA2_LEIQH	P46111 leiurus qui
30	27	26.0	38	1	SCA3_LEIQH	P46112 leiurus qui
31	27	26.0	38	1	SCK1_ANDMA	P24662 androctonus
32	27	26.0	38	1	SCK1_ORTSC	P55896 orthochirus
33	27	26.0	43	1	DEFA_ZOPAT	P80033 zophobas at
34	27	26.0	44	1	IAA3_WHEAT	P10846 triticum ae
35	27	26.0	50	1	ORNE_PLAOR	P25514 placobdella
36	26.5	25.5	39	1	TX4K_EURCA	P18928 eurypelma c
37	26.5	25.5	39	1	TXP1_BRASM	P49265 brachypelma
38	26.5	25.5	49	1	TX25_PHONI	P29424 phoneutria
39	26	25.0	19	1	CXA2_CONST	P28879 conus stria
40	26	25.0	23	1	PAP1_MANSE	P30253 manduca sex
41	26	25.0	27	1	MT2_COLGL	Q00369 colletotric
42	26	25.0	27	1	TXA3_ANESU	P01535 anemonia su
43	26	25.0	35	1	TXKS_STOHE	P29187 stoichactis
44	26	25.0	39	1	DECO_MACDE	P17350 macrobdella
45	26	25.0	43	1	MUTI_ENTMU	P80925 enterococcu

ALIGNMENTS

RESULT 1
THN_DENCL
ID THN_DENCL STANDARD; PRT; 46 AA.
AC P01541;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1988 (Rel. 09, Last annotation update)
DE Denclatoxin B.
OS Dendrophthora clavata (Columbian mistletoe).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Santalales; Viscaceae; Dendrophthora.
OX NCBI_TaxID=3965;
RN [1]
RP SEQUENCE.
RX MEDLINE=78016835; PubMed=906843;
RA Samuelsson G., Pettersson B.;
RT "Toxic proteins from the mistletoe Dendrophtora clavata. II. The
RT amino acid sequence of denclatoxin B.";
RL Acta Pharm. Suec. 14:245-254(1977).
CC -!- FUNCTION: THIONINS ARE SMALL PLANT PROTEINS WHICH ARE TOXIC
CC TO ANIMAL CELLS. THEY SEEM TO EXERT THEIR TOXIC EFFECT AT THE
CC LEVEL OF THE CELL MEMBRANE. THE PRECISE FUNCTION, IN PLANTS,
CC OF THESE PROTEINS IS NOT KNOWN.
CC -!- SIMILARITY: BELONGS TO THE PLANT THIONIN FAMILY.
DR PIR; A01804; DKDCB.
DR HSSP; P01542; 1CBN.
DR InterPro; IPR001010; Thionin.
DR Pfam; PF00321; plant_thionins; 1.
DR PRINTS; PR00287; THIONIN.
DR PROSITE; PS00271; THIONIN; 1.
KW Thionin; Plant toxin.
FT DISULFID 3 40 BY SIMILARITY.
FT DISULFID 4 32 BY SIMILARITY.
FT DISULFID 16 26 BY SIMILARITY.
SQ SEQUENCE 46 AA; 4821 MW; C107A82B29ADA608 CRC64;

Query Match 34.6%; Score 36; DB 1; Length 46;
Best Local Similarity 66.7%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      10 SXXGCRPGY 18
           |  || |||
Db      36 SGTGCPPGY 44
```

RESULT 2
CP23_SPOER
ID CP23_SPOER STANDARD; PRT; 23 AA.
AC P56683;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Cardioactive peptide CAP23.
OS Spodoptera eridania (Southern armyworm).

```
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
OX   NCBI_TaxID=37547;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=99196260; PubMed=10098624;
RA   Furuya K., Hackett M., Cirelli M.A., Schegg K.M., Wang H.,
RA   Shabanowitz J., Hunt D.F., Schooley D.A.;
RT   "A cardioactive peptide from the southern armyworm, Spodoptera
RT   eridania.";
RL   Peptides 20:53-61(1999).
CC   -!- FUNCTION: HAS EXCITATORY EFFECTS ON A SEMI-ISOLATED HEART FROM
CC       LARVAL MANDUCA SEXTA, CAUSING AN INOTROPIC EFFECT AT LOW
CC       CONCENTRATIONS OF PEPTIDE AND CHRONOTROPIC AND INOTROPIC EFFECTS
CC       AT HIGH DOSES.
CC   -!- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
DR   HSSP; O61704; 1B5N.
DR   InterPro; IPR003463; GBP_PSP.
DR   Pfam; PF02425; GBP_PSP; 1.
FT   DISULFID      7     19       BY SIMILARITY.
SQ   SEQUENCE   23 AA;  2519 MW;  0A96D72A70855AE0 CRC64;


  Query Match             32.7%;  Score 34;  DB 1;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   3
PAP1_SPOEX
ID   PAP1_SPOEX     STANDARD;      PRT;    23 AA.
AC   P30255;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Paralytic peptide I (PP I).
OS   Spodoptera exigua (Beet armyworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
OX   NCBI_TaxID=7107;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Hemolymph;
RX   MEDLINE=91302298; PubMed=2071576;
RA   Skinner W.S., Dennis P.A., Li J.P., Summerfelt R.M., Carney R.L.,
RA   Quistad G.B.;
RT   "Isolation and identification of paralytic peptides from hemolymph of
RT   the lepidopteran insects Manduca sexta, Spodoptera exigua, and
RT   Heliothis virescens.";
RL   J. Biol. Chem. 266:12873-12877(1991).
CC   -!- FUNCTION: CAUSES RAPID, RIGID PARALYSIS WHEN INJECTED INTO
CC       LEPIDOPTERAN LARVAE. THE PHYSIOLOGICAL ROLE MAY BE TO REDUCE
CC       HEMOLYMPH LOSS FOLLOWING INJURY AND PROMOTE WOUND HEALING.
CC   -!- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
DR   PIR; C39855; C39855.
DR   HSSP; O61704; 1B5N.
DR   InterPro; IPR003463; GBP_PSP.
DR   Pfam; PF02425; GBP_PSP; 1.
KW   Hemolymph.
FT   DISULFID      7     19       BY SIMILARITY.
SQ   SEQUENCE   23 AA;  2451 MW;  0A96D1F600855AE0 CRC64;


  Query Match             32.7%;  Score 34;  DB 1;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   4
PAP2_SPOEX
ID   PAP2_SPOEX     STANDARD;      PRT;    23 AA.
AC   P30256;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Paralytic peptide II (PP II).
OS   Spodoptera exigua (Beet armyworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
OX   NCBI_TaxID=7107;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Hemolymph;
RX   MEDLINE=91302298; PubMed=2071576;
RA   Skinner W.S., Dennis P.A., Li J.P., Summerfelt R.M., Carney R.L.,
RA   Quistad G.B.;
RT   "Isolation and identification of paralytic peptides from hemolymph of
RT   the lepidopteran insects Manduca sexta, Spodoptera exigua, and
RT   Heliothis virescens.";
RL   J. Biol. Chem. 266:12873-12877(1991).
CC   -!- FUNCTION: CAUSES RAPID, RIGID PARALYSIS WHEN INJECTED INTO
CC       LEPIDOPTERAN LARVAE. THE PHYSIOLOGICAL ROLE MAY BE TO REDUCE
CC       HEMOLYMPH LOSS FOLLOWING INJURY AND PROMOTE WOUND HEALING.
CC   -!- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
DR   PIR; D39855; D39855.
DR   HSSP; O61704; 1B5N.
DR   InterPro; IPR003463; GBP_PSP.
DR   Pfam; PF02425; GBP_PSP; 1.
KW   Hemolymph.
FT   DISULFID      7     19       BY SIMILARITY.
SQ   SEQUENCE   23 AA;  2477 MW;  0A96CB4600855AE0 CRC64;


  Query Match             32.7%;  Score 34;  DB 1;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   5
PAP3_SPOEX
ID   PAP3_SPOEX     STANDARD;      PRT;    23 AA.
AC   P30257;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Paralytic peptide III (PP III).
OS   Spodoptera exigua (Beet armyworm).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Amphipyrinae; Spodoptera.
OX   NCBI_TaxID=7107;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Hemolymph;
RX   MEDLINE=91302298; PubMed=2071576;
RA   Skinner W.S., Dennis P.A., Li J.P., Summerfelt R.M., Carney R.L.,
RA   Quistad G.B.;
RT   "Isolation and identification of paralytic peptides from hemolymph of
RT   the lepidopteran insects Manduca sexta, Spodoptera exigua, and
RT   Heliothis virescens.";
RL   J. Biol. Chem. 266:12873-12877(1991).
```

```
CC   -!- FUNCTION: CAUSES RAPID, RIGID PARALYSIS WHEN INJECTED INTO
CC       LEPIDOPTERAN LARVAE. THE PHYSIOLOGICAL ROLE MAY BE TO REDUCE
CC       HEMOLYMPH LOSS FOLLOWING INJURY AND PROMOTE WOUND HEALING.
CC   -!- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
DR   PIR; E39855; E39855.
DR   HSSP; O61704; 1B5N.
DR   InterPro; IPR003463; GBP_PSP.
DR   Pfam; PF02425; GBP_PSP; 1.
KW   Hemolymph.
FT   DISULFID      7     19       BY SIMILARITY.
SQ   SEQUENCE   23 AA;  2505 MW;  0A96CB5EB7D55AE0 CRC64;


  Query Match             32.7%;  Score 34;  DB 1;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 21;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCTPGY 11


RESULT   6
CXA4_CONPU
ID   CXA4_CONPU     STANDARD;      PRT;    25 AA.
AC   P55963;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Alpha-A conotoxin PIVA.
OS   Conus purpurascens (Purple cone).
OC   Eukaryota; Metazoa; Mollusca; Gastropoda; Caenogastropoda;
OC   Neogastropoda; Conoidea; Conidae; Conus.
OX   NCBI_TaxID=41690;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=95403432; PubMed=7673220;
RA   Hopkins C., Grilley M., Miller C., Shon K.-J., Cruz L.J., Gray W.R.,
RA   Dykert J., Rivier J., Yoshikami D., Olivera B.M.;
RT   "A new family of Conus peptides targeted to the nicotinic
RT   acetylcholine receptor.";
RL   J. Biol. Chem. 270:22361-22367(1995).
RN   [2]
RP   STRUCTURE BY NMR.
RX   MEDLINE=97200721; PubMed=9048550;
RA   Han K.-H., Hwang K.-J., Kim S.-M., Kim S.-K., Gray W.R., Olivera B.M.,
RA   Rivier J., Shon K.-J.;
RT   "NMR structure determination of a novel conotoxin, [Pro 7,13] alpha
RT   A-conotoxin PIVA.";
RL   Biochemistry 36:1669-1677(1997).
CC   -!- FUNCTION: ALPHA-CONOTOXINS ACT ON POSTSYNAPTIC MEMBRANES, THEY
CC       BIND TO THE NICOTINIC ACETYLCHOLINE RECEPTORS (NACHR) AND THUS
CC       INHIBIT THEM.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE ALPHA-TYPE CONOTOXIN FAMILY.
DR   PDB; 1P1P; 07-JUL-97.
KW   Postsynaptic neurotoxin; Acetylcholine receptor inhibitor; Amidation;
KW   Hydroxylation; Venom; 3D-structure.
FT   DISULFID      2     16
FT   DISULFID      3     11
FT   DISULFID     14     23
FT   MOD_RES       7      7       HYDROXYLATION.
FT   MOD_RES      13     13       HYDROXYLATION.
FT   MOD_RES      20     20       HYDROXYLATION.
FT   MOD_RES      25     25       AMIDATION.
SQ   SEQUENCE   25 AA;  2608 MW;  9E2147898D697640 CRC64;


  Query Match             32.2%;  Score 33.5;  DB 1;  Length 25;
  Best Local Similarity   57.1%;  Pred. No. 27;
  Matches    8;  Conservative    0;  Mismatches    1;  Indels    5;  Gaps    2;

Qy       7 HPCSXXGC--RPGY 18
           ||||   |  || |
Db      12 HPCS---CKDRPSY 22


RESULT   7
PAP1_HELVI
ID   PAP1_HELVI     STANDARD;      PRT;    23 AA.
AC   P30251;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Paralytic peptide I (PP I).
OS   Heliothis virescens (Noctuid moth) (Owlet moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Heliothinae; Heliothis.
OX   NCBI_TaxID=7102;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Hemolymph;
RX   MEDLINE=91302298; PubMed=2071576;
RA   Skinner W.S., Dennis P.A., Li J.P., Summerfelt R.M., Carney R.L.,
RA   Quistad G.B.;
RT   "Isolation and identification of paralytic peptides from hemolymph of
RT   the lepidopteran insects Manduca sexta, Spodoptera exigua, and
RT   Heliothis virescens.";
RL   J. Biol. Chem. 266:12873-12877(1991).
CC   -!- FUNCTION: CAUSES RAPID, RIGID PARALYSIS WHEN INJECTED INTO
CC       LEPIDOPTERAN LARVAE. THE PHYSIOLOGICAL ROLE MAY BE TO REDUCE
CC       HEMOLYMPH LOSS FOLLOWING INJURY AND PROMOTE WOUND HEALING.
CC   -!- SIMILARITY: BELONGS TO THE GBP / PSP1 / PARALYTIC PEPTIDE FAMILY.
DR   PIR; F39855; F39855.
DR   HSSP; O61704; 1B5N.
DR   InterPro; IPR003463; GBP_PSP.
DR   Pfam; PF02425; GBP_PSP; 1.
KW   Hemolymph.
FT   DISULFID      7     19       BY SIMILARITY.
SQ   SEQUENCE   23 AA;  2524 MW;  2236CB436D655AFA CRC64;


  Query Match             30.8%;  Score 32;  DB 1;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 43;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCIPGY 11


RESULT   8
PAP2_HELVI
ID   PAP2_HELVI     STANDARD;      PRT;    23 AA.
AC   P30252;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-APR-1993 (Rel. 25, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   Paralytic peptide II (PP II).
OS   Heliothis virescens (Noctuid moth) (Owlet moth).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC   Noctuoidea; Noctuidae; Heliothinae; Heliothis.
OX   NCBI_TaxID=7102;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Hemolymph;
RX   MEDLINE=91302298; PubMed=2071576;
RA   Skinner W.S., Dennis P.A., Li J.P., Summerfelt R.M., Carney R.L.,
RA   Quistad G.B.;
RT   "Isolation and identification of paralytic peptides from hemolymph of
RT   the lepidopteran insects Manduca sexta, Spodoptera exigua, and
RT   Heliothis virescens.";
RL   J. Biol. Chem. 266:12873-12877(1991).
```

```
CC   -!- FUNCTION: CAUSES RAPID, RIGID PARALYSIS WHEN INJECTED INTO
CC       LEPIDOPTERAN LARVAE. THE PHYSIOLOGICAL ROLE MAY BE TO REDUCE
CC       HEMOLYMPH LOSS FOLLOWING INJURY AND PROMOTE WOUND HEALING.
DR   PIR; G39855; G39855.
DR   HSSP; O61704; 1B5N.
DR   InterPro; IPR003463; GBP_PSP.
DR   Pfam; PF02425; GBP_PSP; 1.
KW   Hemolymph.
SQ   SEQUENCE   23 AA;  2508 MW;  2236CB5D6C855AFA CRC64;


  Query Match             30.8%;  Score 32;  DB 1;  Length 23;
  Best Local Similarity   83.3%;  Pred. No. 43;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     13 GCRPGY 18
          || |||
Db      6 GCIPGY 11


RESULT   9
DCMS_PSECA
ID   DCMS_PSECA     STANDARD;      PRT;    21 AA.
AC   P19921;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   01-JUN-1994 (Rel. 29, Last annotation update)
DE   Carbon monoxide oxygenase [cytochrome B-561] small chain (EC 1.2.2.4)
DE   (Fragment).
OS   Pseudomonas carboxydovorans.
OC   Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC   Bradyrhizobium group; Oligotropha.
OX   NCBI_TaxID=40137;
RN   [1]
RP   SEQUENCE.
RC   STRAIN=OM5;
RX   MEDLINE=90055678; PubMed=2818128;
RA   Kraut M., Hugendieck I., Herwig S., Meyer O.;
RT   "Homology and distribution of CO dehydrogenase structural genes in
RT   carboxydotrophic bacteria.";
RL   Arch. Microbiol. 152:335-341(1989).
CC   -!- CATALYTIC ACTIVITY: CO + H(2)O + ferrocytochrome b-561 = CO(2) + 2
CC       H(+) + ferricytochrome b-561.
CC   -!- COFACTOR: MOLYBDENUM.
CC   -!- SUBUNIT: CONSISTS OF THREE POLYPEPTIDE CHAINS: LARGE, MEDIUM, AND
CC       SMALL.
DR   PIR; PL0144; PL0144.
KW   Oxidoreductase; Molybdenum.
FT   NON_TER      21     21
SQ   SEQUENCE   21 AA;  2270 MW;  68D4380629401B9C CRC64;


  Query Match             29.8%;  Score 31;  DB 1;  Length 21;
  Best Local Similarity   83.3%;  Pred. No. 57;
  Matches    5;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      3 TIHGHP 8
          ||:|||
Db      9 TINGHP 14


RESULT  10
VARB_VIOAR
ID   VARB_VIOAR     STANDARD;      PRT;    30 AA.
AC   P58447;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Varv peptide B.
OS   Viola arvensis (European field pansy) (Field violet).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Malpighiales; Violaceae; Viola.
OX   NCBI_TaxID=97415;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=99177275; PubMed=10075760;
RA   Goeransson U., Luijendijk T., Johansson S., Bohlin L., Claeson P.;
RT   "Seven novel macrocyclic polypeptides from Viola arvensis.";
RL   J. Nat. Prod. 62:283-286(1999).
CC   -!- FUNCTION: Probably participates in a plant defense mechanism.
CC   -!- PTM: This is a cyclic peptide.
CC   -!- CAUTION: This peptide is cyclic, its sequence was chosen to start
CC       at the position shown below by similarity to Oak1 (kalata B1)
CC       whose DNA sequence is known.
KW   Multigene family.
FT   DISULFID      5     19
FT   DISULFID      9     21
FT   DISULFID     14     27
SQ   SEQUENCE   30 AA;  3093 MW;  7B09691FEBAD26EE CRC64;


  Query Match             29.8%;  Score 31;  DB 1;  Length 30;
  Best Local Similarity   38.5%;  Pred. No. 79;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 DTIHGHPCSXXGC 14
          :|  |  |:  ||
Db      7 ETCFGGTCNTPGC 19


RESULT  11
VARG_VIOAR
ID   VARG_VIOAR     STANDARD;      PRT;    30 AA.
AC   P58452;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Varv peptide G.
OS   Viola arvensis (European field pansy) (Field violet).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Malpighiales; Violaceae; Viola.
OX   NCBI_TaxID=97415;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=99177275; PubMed=10075760;
RA   Goeransson U., Luijendijk T., Johansson S., Bohlin L., Claeson P.;
RT   "Seven novel macrocyclic polypeptides from Viola arvensis.";
RL   J. Nat. Prod. 62:283-286(1999).
CC   -!- FUNCTION: Probably participates in a plant defense mechanism.
CC   -!- PTM: This is a cyclic peptide.
CC   -!- CAUTION: This peptide is cyclic, its sequence was chosen to start
CC       at the position shown below by similarity to Oak1 (kalata B1)
CC       whose DNA sequence is known.
KW   Multigene family.
FT   DISULFID      5     19
FT   DISULFID      9     21
FT   DISULFID     14     27
SQ   SEQUENCE   30 AA;  3047 MW;  7B09691FE45C9CEE CRC64;


  Query Match             29.8%;  Score 31;  DB 1;  Length 30;
  Best Local Similarity   38.5%;  Pred. No. 79;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      2 DTIHGHPCSXXGC 14
          :|  |  |:  ||
Db      7 ETCFGGTCNTPGC 19


RESULT  12
VARH_VIOAR
ID   VARH_VIOAR     STANDARD;      PRT;    30 AA.
AC   P58453;
DT   01-MAR-2002 (Rel. 41, Created)
```

```
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Varv peptide H.
OS   Viola arvensis (European field pansy) (Field violet).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Malpighiales; Violaceae; Viola.
OX   NCBI_TaxID=97415;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=99177275; PubMed=10075760;
RA   Goeransson U., Luijendijk T., Johansson S., Bohlin L., Claeson P.;
RT   "Seven novel macrocyclic polypeptides from Viola arvensis.";
RL   J. Nat. Prod. 62:283-286(1999).
CC   -!- FUNCTION: Probably participates in a plant defense mechanism.
CC   -!- PTM: This is a cyclic peptide.
CC   -!- CAUTION: This peptide is cyclic, its sequence was chosen to start
CC       at the position shown below by similarity to Oak1 (kalata B1)
CC       whose DNA sequence is known.
KW   Multigene family.
FT   DISULFID      5     19
FT   DISULFID      9     21
FT   DISULFID     14     27
SQ   SEQUENCE   30 AA;  3079 MW;  CE4C691FFFFD26E8 CRC64;


  Query Match             29.8%;  Score 31;  DB 1;  Length 30;
  Best Local Similarity   38.5%;  Pred. No. 79;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       2 DTIHGHPCSXXGC 14
           :|  |  |:  ||
Db       7 ETCFGGTCNTPGC 19


RESULT  13
VG38_BPMD2
ID   VG38_BPMD2     STANDARD;      PRT;    50 AA.
AC   O64229;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Gene 38 protein (GP38).
GN   38.
OS   Mycobacteriophage D29.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=28369;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98300335; PubMed=9636706;
RA   Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT   "Genome structure of mycobacteriophage D29: implications for phage
RT   evolution.";
RL   J. Mol. Biol. 279:143-164(1998).
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AF022214; AAC18479.1; -.
SQ   SEQUENCE   50 AA;  4851 MW;  75BCC1A1CF2EF26E CRC64;


  Query Match             29.8%;  Score 31;  DB 1;  Length 50;
  Best Local Similarity   50.0%;  Pred. No. 1.3e+02;
  Matches    5;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       9 CSXXGCRPGY 18
           |   |  |||
Db      20 CDGGGSAPGY 29


RESULT  14
SCX6_TITBA
ID   SCX6_TITBA     STANDARD;      PRT;    19 AA.
AC   P56610;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Toxin TbTX-VI (Fragment).
OS   Tityus bahiensis (Brazilian scorpion).
OC   Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Scorpiones;
OC   Buthoidea; Buthidae; Tityus.
OX   NCBI_TaxID=50343;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=Venom;
RX   MEDLINE=96190713; PubMed=8611151;
RA   Becerril B., Corona M., Coronas F.I., Zamudio F.,
RA   Calderon-Aranda E.S., Fletcher P.L. Jr., Martin B.M., Possani L.D.;
RT   "Toxic peptides and genes encoding toxin gamma of the Brazilian
RT   scorpions Tityus bahiensis and Tityus stigmurus.";
RL   Biochem. J. 313:753-760(1996).
CC   -!- FUNCTION: NOT TOXIC IN MICE.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: BELONGS TO THE ALPHA/BETA-SCORPION TOXIN FAMILY.
CC       ALPHA-TOXIN SUBFAMILY.
FT   NON_TER      19     19
SQ   SEQUENCE   19 AA;  2151 MW;  3535A2F1E5E67D14 CRC64;


  Query Match             27.9%;  Score 29;  DB 1;  Length 19;
  Best Local Similarity   40.0%;  Pred. No. 1.1e+02;
  Matches    4;  Conservative    2;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       6 GHPCSXXGCR 15
           |:|    ||:
Db       4 GYPTDKRGCK 13


RESULT  15
CYOC_VIOOD
ID   CYOC_VIOOD     STANDARD;      PRT;    29 AA.
AC   P58444;
DT   01-MAR-2002 (Rel. 41, Created)
DT   01-MAR-2002 (Rel. 41, Last sequence update)
DT   01-MAR-2002 (Rel. 41, Last annotation update)
DE   Cycloviolacin O12.
OS   Viola odorata (Sweet violet).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids I; Malpighiales; Violaceae; Viola.
OX   NCBI_TaxID=97441;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=20069951; PubMed=10600388;
RA   Craik D.J., Daly N.L., Bond T., Waine C.;
RT   "Plant cyclotides: a unique family of cyclic and knotted proteins that
RT   defines the cyclic cystine knot structural motif.";
RL   J. Mol. Biol. 294:1327-1336(1999).
CC   -!- FUNCTION: Probably participates in a plant defense mechanism.
CC   -!- PTM: This is a cyclic peptide.
CC   -!- CAUTION: This peptide is cyclic, its sequence was chosen to start
CC       at the position shown below by similarity to Oak1 (kalata B1)
CC       whose DNA sequence is known.
KW   Multigene family.
FT   DISULFID      5     19
FT   DISULFID      9     21
FT   DISULFID     14     26
SQ   SEQUENCE   29 AA;  2916 MW;  323641013F82FA18 CRC64;
```

Query Match 27.9%; Score 29; DB 1; Length 29;
Best Local Similarity 38.5%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```
Qy       2 DTIHGHPCSXXGC 14
           :|  |  |:  ||
Db       7 ETCVGGTCNTPGC 19
```

Search completed: August 26, 2002, 13:43:29
Job time: 353 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:42:51 ; Search time 35.18 Seconds
(without alignments)
88.514 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 29986

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	36	34.6	31	13	P82878	P82878 rana clamit
2	36	34.6	32	2	O05602	O05602 pseudomonas
3	36	34.6	36	11	Q9JMC0	Q9jmc0 rattus norv
4	35.5	34.1	42	6	O18958	O18958 bos taurus
5	34	32.7	27	13	P82879	P82879 rana clamit
6	34	32.7	47	6	O97978	O97978 equus cabal
7	34	32.7	47	6	O97977	O97977 equus cabal
8	34	32.7	47	6	Q9N1F7	Q9n1f7 equus cabal
9	33	31.7	36	12	Q91D77	Q91d77 ttv-like mi
10	33	31.7	39	15	O36981	O36981 caprine art
11	33	31.7	41	12	Q91D79	Q91d79 ttv-like mi
12	33	31.7	41	12	Q91D78	Q91d78 ttv-like mi
13	33	31.7	46	3	Q9HFA8	Q9hfa8 trichosporo
14	33	31.7	47	2	Q9F3V1	Q9f3v1 pseudonocar
15	32	30.8	22	4	Q9Y6S3	Q9y6s3 homo sapien
16	32	30.8	26	3	O93940	O93940 podospora a
17	32	30.8	44	11	Q920R8	Q920r8 mus musculu
18	31.5	30.3	37	11	Q9QX87	Q9qx87 rattus norv
19	31	29.8	37	2	Q9KHN7	Q9khn7 pseudomonas
20	31	29.8	37	2	Q9KHN5	Q9khn5 pseudomonas
21	31	29.8	38	15	Q66720	Q66720 equine infe
22	30.5	29.3	38	4	Q9UGU2	Q9ugu2 homo sapien
23	30	28.8	19	8	O63058	O63058 lathraea cl
24	30	28.8	37	5	Q9U678	Q9u678 crypthecodi
25	30	28.8	41	12	Q91U54	Q91u54 influenza a
26	30	28.8	43	12	Q91U58	Q91u58 influenza a
27	30	28.8	43	12	Q91U52	Q91u52 influenza a
28	30	28.8	44	15	Q993M8	Q993m8 human immun
29	30	28.8	47	16	Q9JXF4	Q9jxf4 neisseria m
30	29.5	28.4	30	10	Q96240	Q96240 arabidopsis
31	29.5	28.4	37	4	Q9H499	Q9h499 homo sapien
32	29	27.9	27	4	Q9NTX3	Q9ntx3 homo sapien
33	29	27.9	31	13	P82742	P82742 rana catesb
34	29	27.9	34	4	Q16125	Q16125 homo sapien
35	29	27.9	36	5	O18611	O18611 anemonia su
36	29	27.9	39	4	Q9H4W0	Q9h4w0 homo sapien
37	29	27.9	43	5	Q9TX99	Q9tx99 caenorhabdi
38	29	27.9	45	4	Q96P40	Q96p40 homo sapien
39	28.5	27.4	48	4	Q9BY82	Q9by82 homo sapien
40	28	26.9	14	4	Q16332	Q16332 homo sapien
41	28	26.9	20	5	Q9GPU4	Q9gpu4 sterkiella
42	28	26.9	27	12	Q68059	Q68059 hepatitis b
43	28	26.9	30	2	Q9Z495	Q9z495 acinetobact
44	28	26.9	36	2	Q9KHN3	Q9khn3 pseudomonas
45	28	26.9	36	2	Q9KHN1	Q9khn1 pseudomonas

ALIGNMENTS

RESULT 1
P82878
ID P82878 PRELIMINARY; PRT; 31 AA.
AC P82878;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE RANATUERIN-2CA.
OS Rana clamitans (green frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX NCBI_TaxID=145282;
RN [1]
RP SEQUENCE.
RC TISSUE=SKIN;
RX MEDLINE=20283865; PubMed=10822101;
RA Halverson T., Basir Y.J., Knoop F.C., Conlon J.M.;
RT "Purification and characterization of antimicrobial peptides from the
RT skin of the North American green frog Rana clamitans.";
RL Peptides 21:469-476(2000).
CC -!- FUNCTION: ANTIBACTERIAL ACTIVITY AGAINST GRAM-POSITIVE BACTERIUM
CC S.AUREUS AND GRAM-NEGATIVE BACTERIUM E.COLI. HAS ACTIVITY AGAINST
CC C.ALBICANS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- MASS SPECTROMETRY: MW=3156.4; MW_ERR=0.02; METHOD=ELECTROSPRAY.
CC -!- SIMILARITY: BELONGS TO THE BREVININ/ESCULENTIN/GAEGURIN/RUGOSIN
CC FAMILY.
KW Antibiotic; Fungicide.
FT DISULFID 24 29
SQ SEQUENCE 31 AA; 3159 MW; 79DC2D856D32D2E0 CRC64;

Query Match 34.6%; Score 36; DB 13; Length 31;
Best Local Similarity 38.5%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

Qy 4 IHGHPCSXXGCRP 16
: | | || :|
Db 19 LEGLKCKIAGCKP 31

```
RESULT   2
O05602
ID   O05602      PRELIMINARY;      PRT;    32 AA.
AC   O05602;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TRANSPOSON TN5041 DNA (FRAGMENT).
OS   Pseudomonas sp.
OC   Bacteria; Proteobacteria.
OX   NCBI_TaxID=306;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KHP41; TRANSPOSON=TN5041;
RX   MEDLINE=97419493; PubMed=9274008;
RA   Khloldii G.Y., Yurieva O.V., Gorlenko Z.M., Mindlin S.Z., Bass I.A.,
RA   Lomovskaya O.L., Kopteva A.V., Nikiforov V.G.;
RT   "Tn5041 : a chimeric mercury resistance transposon closely related to
RT   the toluene degradative transposon Tn4651.";
RL   Microbiology 143:2549-2556(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KHP41; TRANSPOSON=TN5041;
RA   Kholodii G.Y., Mindlin S.Z., Gorlenko Z.M., Bass I.A., Kalyaeva E.S.,
RA   Nikiforov V.;
RT   "Host-dependent transposition of Tn5041.";
RL   Russ. J. Genet. 36:365-373(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=KHP41; TRANSPOSON=TN5041;
RA   Kholodii G.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X98999; CAA67458.1; -.
FT   NON_TER       1      1
FT   NON_TER      32     32
SQ   SEQUENCE   32 AA;  3298 MW;  AF42B5EEF917077A CRC64;

  Query Match             34.6%;  Score 36;  DB 2;  Length 32;
  Best Local Similarity   66.7%;  Pred. No. 29;
  Matches    6;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       8 PCSXXGCRP 16
           | |  ||||
Db      21 PSSAYGCRP 29


RESULT   3
Q9JMC0
ID   Q9JMC0      PRELIMINARY;      PRT;    36 AA.
AC   Q9JMC0;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE   NEURTURIN (FRAGMENT).
GN   NTN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SD; TISSUE=URINARY BLADDER;
RA   Kawakami T., Wakabayashi Y., Aimi Y., Isono T., Okada Y.;
RT   "Developmental expression of neurturin and GDNF in rat urinary
RT   bladder.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB032562; BAA92851.1; -.
FT   NON_TER       1      1
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4220 MW;  FB8E6826FF31354F CRC64;

  Query Match             34.6%;  Score 36;  DB 11;  Length 36;
  Best Local Similarity   46.2%;  Pred. No. 33;
  Matches    6;  Conservative    1;  Mismatches    2;  Indels    4;  Gaps    1;

Qy       4 IHGHPCSXXGCRP 16
           :  |||    |||
Db       1 VRAHPC----CRP 9


RESULT   4
O18958
ID   O18958      PRELIMINARY;      PRT;    42 AA.
AC   O18958;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   BONE MORPHOGENETIC PROTEIN 1 (FRAGMENT).
GN   BMP1.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98266882; PubMed=9605845;
RA   Martin-Burriel I., Goldammer T., Elduque C., Lunidin M., Barendse W.,
RA   Zaragoza P., Olsaker I.;
RT   "Physical and linkage mapping of the Bovine bone morphogenetic protein
RT   1 on an evolutionary breakpoint of BTA8.";
RL   Cytogenet. Cell Genet. 79:179-183(1997).
DR   EMBL; Y14605; CAA74948.1; -.
DR   HSSP; P00736; 1APQ.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   Pfam; PF00008; EGF; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
KW   Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
FT   NON_TER       1      1
FT   NON_TER      42     42
SQ   SEQUENCE   42 AA;  4739 MW;  4E5967160BCF9B24 CRC64;

  Query Match             34.1%;  Score 35.5;  DB 6;  Length 42;
  Best Local Similarity   41.2%;  Pred. No. 47;
  Matches    7;  Conservative    3;  Mismatches    4;  Indels    3;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPGY 18
           :|:  : ||   | |||
Db      19 NTLGSYKCS---CDPGY 32


RESULT   5
P82879
ID   P82879      PRELIMINARY;      PRT;    27 AA.
AC   P82879;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   RANATUERIN-2CB.
OS   Rana clamitans (green frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Neobatrachia; Ranoidea; Ranidae; Rana.
OX   NCBI_TaxID=145282;
RN   [1]
RP   SEQUENCE.
RC   TISSUE=SKIN;
```

```
RX   MEDLINE=20283865; PubMed=10822101;
RA   Halverson T., Basir Y.J., Knoop F.C., Conlon J.M.;
RT   "Purification and characterization of antimicrobial peptides from the
RT   skin of the North American green frog Rana clamitans.";
RL   Peptides 21:469-476(2000).
CC   -!- FUNCTION: ANTIBACTERIAL ACTIVITY AGAINST GRAM-POSITIVE BACTERIUM
CC       S.AUREUS AND GRAM-NEGATIVE BACTERIUM E.COLI. HAS ACTIVITY AGAINST
CC       C.ALBICANS.
CC   -!- SUBCELLULAR LOCATION: SECRETED.
CC   -!- MASS SPECTROMETRY: MW=2784.0; MW_ERR=0.02; METHOD=ELECTROSPRAY.
CC   -!- SIMILARITY: BELONGS TO THE BREVININ/ESCULENTIN/GAEGURIN/RUGOSIN
CC       FAMILY.
KW   Antibiotic; Fungicide.
FT   DISULFID     20     25
SQ   SEQUENCE   27 AA;  2786 MW;  9912DD7904E723A0 CRC64;


  Query Match             32.7%;  Score 34;  DB 13;  Length 27;
  Best Local Similarity   38.5%;  Pred. No. 54;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       4 IHGHPCSXXGCRP 16
           : |  |   ||:|
Db      15 LQGLKCIKAGCKP 27


RESULT   6
O97978
ID   O97978      PRELIMINARY;      PRT;    47 AA.
AC   O97978;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TRANSFERRIN (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brandon R.B., Giffard J.M., Bell T.K.;
RT   "Single Nucleotide Polymorphisms in Equine Transferrin.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF103829; AAC78349.1; -.
DR   EMBL; AF103826; AAC78346.2; -.
DR   EMBL; AF103827; AAC78347.2; -.
DR   EMBL; AF103828; AAC78348.1; -.
DR   HSSP; P02788; 1CB6.
DR   InterPro; IPR001156; Transferrin.
DR   Pfam; PF00405; transferrin; 1.
FT   NON_TER       1      1
FT   NON_TER      47     47
SQ   SEQUENCE   47 AA;  5278 MW;  BF04EFE460A64228 CRC64;


  Query Match             32.7%;  Score 34;  DB 6;  Length 47;
  Best Local Similarity   83.3%;  Pred. No. 94;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCAPGY 11


RESULT   7
O97977
ID   O97977      PRELIMINARY;      PRT;    47 AA.
AC   O97977;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   TRANSFERRIN (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Brandon R.B., Giffard J.M., Bell T.K.;
RT   "Single Nucleotide Polymorphisms in Equine Transferrin.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF103867; AAC78387.2; -.
DR   EMBL; AF103830; AAC78350.2; -.
DR   EMBL; AF103831; AAC78351.1; -.
DR   EMBL; AF103832; AAC78352.1; -.
DR   EMBL; AF103833; AAC78353.1; -.
DR   EMBL; AF103834; AAC78354.1; -.
DR   HSSP; P02788; 1CB6.
DR   InterPro; IPR001156; Transferrin.
DR   Pfam; PF00405; transferrin; 1.
FT   NON_TER       1      1
FT   NON_TER      47     47
SQ   SEQUENCE   47 AA;  5252 MW;  BF04EFE46A133218 CRC64;


  Query Match             32.7%;  Score 34;  DB 6;  Length 47;
  Best Local Similarity   83.3%;  Pred. No. 94;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCAPGY 11


RESULT   8
Q9N1F7
ID   Q9N1F7      PRELIMINARY;      PRT;    47 AA.
AC   Q9N1F7;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TRANSFERRIN (FRAGMENT).
OS   Equus caballus (Horse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX   NCBI_TaxID=9796;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Giffard J.M., Brandon R.B., Bell T.K.;
RT   "Further identification of single nucleotide polymorphisms in the
RT   equine transferrin gene.";
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF185786; AAF44352.1; -.
DR   HSSP; P02788; 1CB6.
DR   InterPro; IPR001156; Transferrin.
DR   Pfam; PF00405; transferrin; 1.
FT   NON_TER       1      1
FT   NON_TER      47     47
SQ   SEQUENCE   47 AA;  5251 MW;  BF04EFE460B39818 CRC64;


  Query Match             32.7%;  Score 34;  DB 6;  Length 47;
  Best Local Similarity   83.3%;  Pred. No. 94;
  Matches    5;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy      13 GCRPGY 18
           || |||
Db       6 GCAPGY 11


RESULT   9
Q91D77
ID   Q91D77      PRELIMINARY;      PRT;    36 AA.
AC   Q91D77;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF2 HYPOTHETICAL PROTEIN, ISOLATE:HM0319 (FRAGMENT).
OS   TTV-like mini virus.
OC   Viruses; ssDNA viruses; Circoviridae.
OX   NCBI_TaxID=93678;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM0319;
RA   Michitaka K., Matsubara H., Horiike N., Kihana T., Yano M., Mori T.,
RA   Onji M.;
RT   "Existence of TT virus DNA and TTV-like mini virus DNA in infant cord
RT   blood.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB059561; BAB69654.1; -.
KW   Hypothetical protein.
FT   NON_TER      36     36
SQ   SEQUENCE   36 AA;  4291 MW;  92145F475EA841F1 CRC64;

  Query Match           31.7%;  Score 33;  DB 12;  Length 36;
  Best Local Similarity  38.5%;  Pred. No. 1.1e+02;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRP 16
          :|||      |:|
Db     23 VHGHDIFCDCCKP 35


RESULT  10
O36981
ID   O36981      PRELIMINARY;      PRT;    39 AA.
AC   O36981;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TAT (FRAGMENT).
GN   TAT.
OS   Caprine arthritis encephalitis virus (CAEV).
OC   Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX   NCBI_TaxID=11660;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CA680;
RX   MEDLINE=98022957; PubMed=9356342;
RA   Valas S., Benoit C., Guionaud C., Perrin G., Mamoun R.Z.;
RT   "North american and french caprine arthritis-encephalitis viruses
RT   emerge from ovine maedi-visna viruses.";
RL   Virology 237:307-318(1997).
DR   EMBL; AF015180; AAB87043.1; -.
DR   InterPro; IPR004247; Lentiviral_Tat.
DR   Pfam; PF02998; Lentiviral_Tat; 1.
FT   NON_TER      1      1
SQ   SEQUENCE   39 AA;  4678 MW;  86E38912AFCB369A CRC64;

  Query Match           31.7%;  Score 33;  DB 15;  Length 39;
  Best Local Similarity  45.5%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    1;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      8 PCSXXGCRPGY 18
          ||    | ||:
Db     27 PCGCRLCNPGW 37


RESULT  11
Q91D79
ID   Q91D79      PRELIMINARY;      PRT;    41 AA.
AC   Q91D79;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF2 HYPOTHETICAL PROTEIN, ISOLATE:HM0311 (FRAGMENT).
OS   TTV-like mini virus.
OC   Viruses; ssDNA viruses; Circoviridae.
OX   NCBI_TaxID=93678;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM0311;
RA   Michitaka K., Matsubara H., Horiike N., Kihana T., Yano M., Mori T.,
RA   Onji M.;
RT   "Existence of TT virus DNA and TTV-like mini virus DNA in infant cord
RT   blood.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB059559; BAB69652.1; -.
KW   Hypothetical protein.
FT   NON_TER      41     41
SQ   SEQUENCE   41 AA;  5002 MW;  509CDFEE2DF59804 CRC64;

  Query Match           31.7%;  Score 33;  DB 12;  Length 41;
  Best Local Similarity  38.5%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRP 16
          :|||      |:|
Db     28 VHGHDIFCDCCKP 40


RESULT  12
Q91D78
ID   Q91D78      PRELIMINARY;      PRT;    41 AA.
AC   Q91D78;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF2 HYPOTHETICAL PROTEIN, ISOLATE:HM0315 (FRAGMENT).
OS   TTV-like mini virus.
OC   Viruses; ssDNA viruses; Circoviridae.
OX   NCBI_TaxID=93678;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HM0315;
RA   Michitaka K., Matsubara H., Horiike N., Kihana T., Yano M., Mori T.,
RA   Onji M.;
RT   "Existence of TT virus DNA and TTV-like mini virus DNA in infant cord
RT   blood.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB059560; BAB69653.1; -.
KW   Hypothetical protein.
FT   NON_TER      41     41
SQ   SEQUENCE   41 AA;  5002 MW;  262CDFEE2DF59800 CRC64;

  Query Match           31.7%;  Score 33;  DB 12;  Length 41;
  Best Local Similarity  38.5%;  Pred. No. 1.2e+02;
  Matches    5;  Conservative    2;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRP 16
          :|||      |:|
Db     28 VHGHDIFCDCCKP 40


RESULT  13
Q9HFA8
ID   Q9HFA8      PRELIMINARY;      PRT;    46 AA.
AC   Q9HFA8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   DIHYDROLIPOAMIDE DEHYDROGENASE (FRAGMENT).
GN   LPD.
OS   Trichosporon asahii.
OC   Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes;
OC   Tremellomycetidae; Trichosporonales; Trichosporon.
OX   NCBI_TaxID=82508;
RN   [1]
```

```
RP   SEQUENCE FROM N.A.
RA   Usui Y., Matsunaga Y.;
RT   "Trichosporon asahii gene for dihydrolipoamide dehydrogenase.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB038507; BAB20767.1; -.
FT   NON_TER       1       1
SQ   SEQUENCE   46 AA;  4788 MW;  BC4C6B73E93A2B36 CRC64;


  Query Match             31.7%;  Score 33;  DB 3;  Length 46;
  Best Local Similarity   71.4%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 DTIHGHP 8
          || | ||
Db     20 DTCHAHP 26


RESULT  14
Q9F3V1
ID   Q9F3V1      PRELIMINARY;      PRT;    47 AA.
AC   Q9F3V1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE   HYPOTHETICAL 5.1 KDA PROTEIN.
OS   Pseudonocardia sp. K1.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Pseudonocardineae; Pseudonocardiaceae;
OC   Pseudonocardia.
OX   NCBI_TaxID=102884;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RA   Thiemer B., Andreesen J.R., Schraeder T.;
RT   "Molecular analysis of a gene cluster encoding a monooxygenase and a
RT   semialdehyde dehydrogenase involved in tetrahydrofuran degradation by
RT   Pseudonocardia sp. strain K1.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K1;
RA   Thiemer B., Andreesen J.R., Schraeder T.;
RT   "The NADH-dependent reductase of multicomponent tetrahydrofuran
RT   monooxygenase contains a covalently bound FAD.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ296087; CAC10511.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE   47 AA;  5085 MW;  73B14E1F936ABA4C CRC64;


  Query Match             31.7%;  Score 33;  DB 2;  Length 47;
  Best Local Similarity   71.4%;  Pred. No. 1.4e+02;
  Matches    5;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      7 HPCSXXG 13
          ||||  |
Db     15 HPCSRAG 21


RESULT  15
Q9Y6S3
ID   Q9Y6S3      PRELIMINARY;      PRT;    22 AA.
AC   Q9Y6S3;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-NOV-1999 (TrEMBLrel. 12, Last annotation update)
DE   NEUREXIN III (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Young J., Rowen L., Madan A., Qin S., Abbasi N., Dors M., Dahl T.,
RA   Dickhoff R., Hall J., James R., Loretz C., Lasky S., Madan A.,
RA   Prescott S., Ratcliffe A., Shaffer T., Hood L.;
RT   "Sequencing of human chromosome 14 gene for neurexin III.";
RL   Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC007056; AAD41968.1; -.
FT   NON_TER       1       1
FT   NON_TER      22      22
SQ   SEQUENCE   22 AA;  2328 MW;  3821F4BFD125A6C3 CRC64;


  Query Match             30.8%;  Score 32;  DB 4;  Length 22;
  Best Local Similarity   62.5%;  Pred. No. 97;
  Matches    5;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      3 TIHGHPCS 10
          |:| | ||
Db      8 TLHFHSCS 15


Search completed: August 26, 2002, 13:42:52
Job time: 356 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:36:50 ; Search time 40.79 Seconds
(without alignments)
49.015 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 352077

Minimum DB seq length: 0
Maximum DB seq length: 50

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100	96.2	18	22	AAE07225	IGP1650 peptide fo
2	80	76.9	14	22	AAE07227	IGP1676 peptide fo
3	79	76.0	18	22	AAE07221	IGP1646 peptide fo
4	78	75.0	18	22	AAE07220	IGP1611 peptide fo
5	74	71.2	18	22	AAE07222	IGP1647 peptide fo
6	71	68.3	18	22	AAE07223	IGP1648 peptide fo
7	67	64.4	18	22	AAE07224	IGP1649 peptide fo
8	60	57.7	18	22	AAE07230	IGP1685 peptide fo
9	58	55.8	14	22	AAE07226	IGP1651 peptide fo
10	46	44.2	24	22	ABB29263	Peptide #1914 enco
11	46	44.2	24	22	ABB34433	Peptide #1939 enco
12	46	44.2	24	22	ABB19843	Protein #1842 enco
13	46	44.2	24	22	AAM55219	Human brain expres
14	46	44.2	24	22	AAM67615	Human bone marrow
15	46	44.2	24	22	AAM15421	Peptide #1855 enco
16	46	44.2	24	22	AAM03183	Peptide #1865 enco
17	43	41.3	14	22	AAE07231	IGP1686 peptide fo
18	43	41.3	16	21	AAB09455	Hepatitis GB virus
19	42	40.4	17	20	AAY22924	Filagrin derived a
20	42	40.4	17	20	AAY22944	Synthetic peptide
21	42	40.4	17	20	AAY22949	Synthetic peptide
22	42	40.4	18	20	AAY22938	Peptide derived fr
23	42	40.4	18	20	AAY22946	Synthetic peptide
24	42	40.4	30	20	AAY22926	Filagrin derived a
25	42	40.4	47	22	ABB29813	Peptide #2464 enco
26	42	40.4	47	22	ABB34992	Peptide #2498 enco
27	42	40.4	47	22	ABB20400	Protein #2399 enco
28	42	40.4	47	22	AAM55799	Human brain expres
29	42	40.4	47	22	AAM68173	Human bone marrow
30	42	40.4	47	22	AAM15996	Peptide #2430 enco
31	42	40.4	47	22	AAM28496	Peptide #2533 enco
32	42	40.4	47	22	AAM03731	Peptide #2413 enco
33	41	39.4	18	20	AAY22921	Filagrin derived a
34	41	39.4	18	20	AAY22941	Synthetic peptide
35	41	39.4	31	20	AAY22927	Filagrin derived a
36	40	38.5	17	20	AAY22933	Peptide derived fr
37	40	38.5	17	20	AAY22939	Peptide derived fr
38	39	37.5	14	22	AAE07229	IGP1684 peptide fo
39	39	37.5	30	21	AAB39282	Human secreted pro
40	38.5	37.0	49	22	AAE09413	Pea rapid alkalini
41	38.5	37.0	49	22	AAE09414	Alfalfa rapid alka
42	38.5	37.0	49	22	AAE09418	Ice plant rapid al
43	38.5	37.0	49	22	AAE09419	Soybean rapid alka
44	38.5	37.0	49	22	AAE09422	Maize rapid alkali
45	38.5	37.0	49	22	AAE09423	Sorghum rapid alka

ALIGNMENTS

RESULT 1
AAE07225
ID AAE07225 standard; peptide; 18 AA.
XX
AC AAE07225;
XX
DT 06-NOV-2001 (first entry)
XX
DE IGP1650 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW Synthetic peptide; cyclic; IGP1650; autoimmune antibody;
KW rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1..18
FT /note= "Biotinylated residues"
FT Disulfide-bond 9..14
FT Modified-site 11
FT /note= "Citrulline"
FT Modified-site 12
FT /note= "Citrulline"
XX
PN WO200146222-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000WO-EP13037.
XX
PR 21-DEC-1999; 99EP-0870280.
PR 08-SEP-2000; 2000EP-0870195.
XX

```
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1650.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;
```

```
  Query Match             96.2%;  Score 100;  DB 22;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-09;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||||||||||||||||||
Db       1 qdtihghpcsxxgcrpgy 18
```

```
RESULT   2
AAE07227
ID   AAE07227 standard; peptide; 14 AA.
XX
AC   AAE07227;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1676 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1676; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..14
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..14
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
```

```
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1676.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   14 AA;
```

```
  Query Match             76.9%;  Score 80;  DB 22;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 HGHPCSXXGCRPGY 18
           ||||||||||||||
Db       1 hghpcsxxgcrpgy 14
```

```
RESULT   3
AAE07221
ID   AAE07221 standard; peptide; 18 AA.
XX
AC   AAE07221;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1646 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1646; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

PI Union A, Moereels H, Meheus L;
XX
DR WPI; 2001-496657/54.
XX
PT New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT comprises citrulline residue between 2 cysteine residues and is
PT specifically recognized by autoimmune antibodies from patients
PT suffering from rheumatoid arthritis -
XX
PS Claim 9; Page 42; 53pp; English.
XX
CC The present sequence is a cyclic synthetic biotinylated peptide, IGP1646.
CC The peptide comprises a citrulline residue between 2 cysteine residues
CC and is specifically recognised by autoimmune antibodies from patients
CC suffering from rheumatoid arthritis. The peptide comprises amino acids
CC involved in side chain interactions which is essential for the formation
CC of three-dimensional structure of the peptide. The peptide of the
CC invention is useful as a medicament to treat autoimmune diseases,
CC preferably rheumatoid arthritis. It is also useful for treating
CC autoimmune diseases by increasing the size of antigen-immune complexes to
CC improve clearance of the formed immune complexes and for the preparation
CC of a medicament for oral or nasal administration to treat autoimmune
CC diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC to the peptide.
XX
SQ Sequence 18 AA;

Query Match 76.0%; Score 79; DB 22; Length 18;
Best Local Similarity 83.3%; Pred. No. 8e-06;
Matches 15; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy      1 QDTIHGHPCSXXGCRPGY 18
          |||||||||| || | ||
Db      1 qdtihghpcssxghrcgy 18
```

RESULT 4
AAE07220
ID AAE07220 standard; peptide; 18 AA.
XX
AC AAE07220;
XX
DT 06-NOV-2001 (first entry)
XX
DE IGP1611 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW Synthetic peptide; cyclic; IGP1611; autoimmune antibody;
KW rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1..18
FT /note= "Biotinylated residues"
FT Disulfide-bond 9..16
FT Modified-site 11
FT /note= "Citrulline"
FT Modified-site 12
FT /note= "Citrulline"
XX
PN WO200146222-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000WO-EP13037.
XX
PR 21-DEC-1999; 99EP-0870280.
PR 08-SEP-2000; 2000EP-0870195.
XX
PA (INNO-) INNOGENETICS NV.
XX

PI Union A, Moereels H, Meheus L;
XX
DR WPI; 2001-496657/54.
XX
PT New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT comprises citrulline residue between 2 cysteine residues and is
PT specifically recognized by autoimmune antibodies from patients
PT suffering from rheumatoid arthritis -
XX
PS Claim 9; Page 42; 53pp; English.
XX
CC The present sequence is a cyclic synthetic biotinylated peptide, IGP1611.
CC The peptide comprises a citrulline residue between 2 cysteine residues
CC and is specifically recognised by autoimmune antibodies from patients
CC suffering from rheumatoid arthritis. The peptide comprises amino acids
CC involved in side chain interactions which is essential for the formation
CC of three-dimensional structure of the peptide. The peptide of the
CC invention is useful as a medicament to treat autoimmune diseases,
CC preferably rheumatoid arthritis. It is also useful for treating
CC autoimmune diseases by increasing the size of antigen-immune complexes to
CC improve clearance of the formed immune complexes and for the preparation
CC of a medicament for oral or nasal administration to treat autoimmune
CC diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC to the peptide.
XX
SQ Sequence 18 AA;

Query Match 75.0%; Score 78; DB 22; Length 18;
Best Local Similarity 88.9%; Pred. No. 1.2e-05;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1 QDTIHGHPCSXXGCRPGY 18
          |||||||||||| | ||
Db      1 qdtihghpcsxxghrcgy 18
```

RESULT 5
AAE07222
ID AAE07222 standard; peptide; 18 AA.
XX
AC AAE07222;
XX
DT 06-NOV-2001 (first entry)
XX
DE IGP1647 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW Synthetic peptide; cyclic; IGP1647; autoimmune antibody;
KW rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1..18
FT /note= "Biotinylated residues"
FT Disulfide-bond 9..16
FT Modified-site 11
FT /note= "Citrulline"
FT Modified-site 12
FT /note= "Citrulline"
XX
PN WO200146222-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000WO-EP13037.
XX
PR 21-DEC-1999; 99EP-0870280.
PR 08-SEP-2000; 2000EP-0870195.
XX
PA (INNO-) INNOGENETICS NV.
XX

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1647.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             71.2%;  Score 74;  DB 22;  Length 18;
  Best Local Similarity   83.3%;  Pred. No. 4.9e-05;
  Matches   15;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPGY 18
          |||||||||||| : ||
Db      1 qdtihghpcsxxghqcgy 18


RESULT   6
AAE07223
ID   AAE07223 standard; peptide; 18 AA.
XX
AC   AAE07223;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1648 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1648; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1648.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             68.3%;  Score 71;  DB 22;  Length 18;
  Best Local Similarity   88.2%;  Pred. No. 0.00015;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||||||||||| | |
Db      1 qdtihghpcsxxghrcg 17


RESULT   7
AAE07224
ID   AAE07224 standard; peptide; 18 AA.
XX
AC   AAE07224;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1649 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1649; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1649.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;
```

```
  Query Match             64.4%;  Score 67;  DB 22;  Length 18;
  Best Local Similarity   82.4%;  Pred. No. 0.00063;
  Matches   14;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           |||||||||||||  : |
Db       1 qdtihghpcsxxghqcg 17
```

```
RESULT   8
AAE07230
ID   AAE07230 standard; peptide; 18 AA.
XX
AC   AAE07230;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1685 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1685; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..14
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1685.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;
```

```
  Query Match             57.7%;  Score 60;  DB 22;  Length 18;
  Best Local Similarity   70.6%;  Pred. No. 0.008;
  Matches   12;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           ||||  |   |  ||||||
Db       1 qdtivgwgcdsxgcrpg 17
```

```
RESULT   9
AAE07226
ID   AAE07226 standard; peptide; 14 AA.
XX
AC   AAE07226;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1651 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1651; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..14
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
```

DR WPI; 2001-496657/54.
XX
PT New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT comprises citrulline residue between 2 cysteine residues and is
PT specifically recognized by autoimmune antibodies from patients
PT suffering from rheumatoid arthritis -
XX
PS Claim 9; Page 42; 53pp; English.
XX
CC The present sequence is a cyclic synthetic biotinylated peptide, IGP1651.
CC The peptide comprises a citrulline residue between 2 cysteine residues
CC and is specifically recognised by autoimmune antibodies from patients
CC suffering from rheumatoid arthritis. The peptide comprises amino acids
CC involved in side chain interactions which is essential for the formation
CC of three-dimensional structure of the peptide. The peptide of the
CC invention is useful as a medicament to treat autoimmune diseases,
CC preferably rheumatoid arthritis. It is also useful for treating
CC autoimmune diseases by increasing the size of antigen-immune complexes to
CC improve clearance of the formed immune complexes and for the preparation
CC of a medicament for oral or nasal administration to treat autoimmune
CC diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC to the peptide.
XX
SQ Sequence 14 AA;

Query Match 55.8%; Score 58; DB 22; Length 14;
Best Local Similarity 85.7%; Pred. No. 0.013;
Matches 12; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 5 HGHPCSXXGCRPGY 18
||||||||| | ||
Db 1 hghpcsxxghrcgy 14

RESULT 10
ABB29263
ID ABB29263 standard; Peptide; 24 AA.
XX
AC ABB29263;
XX
DT 01-FEB-2002 (first entry)
XX
DE Peptide #1914 encoded by breast cell single exon nucleic acid probe.
XX
KW Human; microarray; single exon probe; gene expression; breast;
KW disease; cancer.
XX
OS Homo sapiens.
XX
PN WO200157271-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00662.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-496933/54.
XX
PT New spatially-addressable set of single exon nucleic acid probes,
PT useful for measuring gene expression in sample derived from human
PT breast, comprises number of single exon nucleic acid probes -
XX
PS Claim 27; SEQ ID NO 12231; 327pp + sequence listing; English.
XX
CC The invention relates to a spatially-addressable set of single exon
CC nucleic acid probes for measuring gene expression in a sample derived
CC from human breast and BT 474 cells. The method involves contacting
CC the probes with a collection of detectably labelled nucleic acids
CC derived from mRNA of human breast, and then measuring the label
CC bound to each probe of the microarray. The probes are useful for
CC verifying the expression of regions of genomic DNA predicted to
CC encode proteins. They are useful for gene discovery, and for
CC determining predisposition and/or prognosing breast disease. Gene
CC expression analysis is useful for assessing the toxicity of chemical
CC agents on cells. The microarray of this invention presents a far greater
CC diversity of probes for measuring gene expression, with far less bias
CC than expressed sequence tag microarrays. The method is suitable for
CC rapid production of functional information from genomic sequence. The
CC present sequence is a peptide encoded by a single exon nucleic acid
CC probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 24 AA;

Query Match 44.2%; Score 46; DB 22; Length 24;
Best Local Similarity 63.6%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 7 HPCSXXGCRPG 17
||| || ||
Db 7 hpccgrgcwpg 17

RESULT 11
ABB34433
ID ABB34433 standard; Peptide; 24 AA.
XX
AC ABB34433;
XX
DT 04-FEB-2002 (first entry)
XX
DE Peptide #1939 encoded by human foetal liver single exon probe.
XX
KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS Homo sapiens.
XX
PN WO200157277-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00669.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483447/52.
XX
PT Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
XX
PS Claim 27; SEQ ID NO 27068; 639pp + sequence listing; English.

XX
CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 24 AA;

Query Match 44.2%; Score 46; DB 22; Length 24;
Best Local Similarity 63.6%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 7 HPCSXXGCRPG 17
||| || ||
Db 7 hpccgrgcwpg 17

RESULT 12
ABB19843
ID ABB19843 standard; Protein; 24 AA.
XX
AC ABB19843;
XX
DT 23-JAN-2002 (first entry)
XX
DE Protein #1842 encoded by probe for measuring heart cell gene expression.
XX
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
XX
OS Homo sapiens.
XX
PN WO200157274-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00666.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488899/53.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT hearts -
XX
PS Claim 15; SEQ ID No 21613; 530pp; English.
XX
CC The present invention relates to single exon nucleic acid probes for
CC measuring human gene expression in a sample derived from human heart (see
CC ABA21535-ABA41305). The present sequence is a protein encoded by one such
CC probe. The probes may be used for predicting, measuring and displaying
CC gene expression in samples derived from the human heart via microarrays.
CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 24 AA;

Query Match 44.2%; Score 46; DB 22; Length 24;
Best Local Similarity 63.6%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 7 HPCSXXGCRPG 17
||| || ||
Db 7 hpccgrgcwpg 17

RESULT 13
AAM55219
ID AAM55219 standard; Protein; 24 AA.
XX
AC AAM55219;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 27324.
XX
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
XX
OS Homo sapiens.
XX
PN WO200157275-A2.
XX
PD 09-AUG-2001.
XX
PF 30-JAN-2001; 2001WO-US00667.
XX
PR 04-FEB-2000; 2000US-0180312.
PR 26-MAY-2000; 2000US-0207456.
PR 30-JUN-2000; 2000US-0608408.
PR 03-AUG-2000; 2000US-0632366.
PR 21-SEP-2000; 2000US-0234687.
PR 27-SEP-2000; 2000US-0236359.
PR 04-OCT-2000; 2000GB-0024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-483446/52.
XX
PT Single exon nucleic acid probes for analyzing gene expression in human
PT brains -
XX
PS Example 4; SEQ ID NO: 27324; 650pp + Sequence Listing; English.
XX
CC The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.
XX
SQ Sequence 24 AA;

Query Match 44.2%; Score 46; DB 22; Length 24;
Best Local Similarity 63.6%; Pred. No. 1.7;
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 7 HPCSXXGCRPG 17

```
          |||   || ||
Db      7 hpccgrgcwpg 17


RESULT  14
AAM67615
ID   AAM67615 standard; Protein; 24 AA.
XX
AC   AAM67615;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human bone marrow expressed probe encoded protein SEQ ID NO: 27921.
XX
KW   Human; bone marrow expressed exon; gene expression analysis; probe;
KW   microarray; cancer; leukaemia; lymphoma; myeloma.
XX
OS   Homo sapiens.
XX
PN   WO200157276-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00668.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488900/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human bone marrow -
XX
PS   Example 4; SEQ ID NO: 27921; 658pp + Sequence Listing; English.
XX
CC   The present invention provides a number of single exon nucleic acid
CC   probes which are derived from genomic sequences expressed in the human
CC   bone marrow. They can be used to measure gene expression in bone marrow
CC   samples, which may enable the improved diagnosis and treatment of cancers
CC   such as lymphoma, leukaemia and myeloma. The present sequence is a
CC   protein encoded by one of the probes of the invention.
XX
SQ   Sequence   24 AA;

  Query Match           44.2%;  Score 46;  DB 22;  Length 24;
  Best Local Similarity   63.6%;  Pred. No. 1.7;
  Matches    7;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 HPCSXXGCRPG 17
          |||   || ||
Db      7 hpccgrgcwpg 17


RESULT  15
AAM15421
ID   AAM15421 standard; Protein; 24 AA.
XX
AC   AAM15421;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Peptide #1855 encoded by probe for measuring cervical gene expression.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US00670.
XX
PR   04-FEB-2000; 2000US-0180312.
PR   26-MAY-2000; 2000US-0207456.
PR   30-JUN-2000; 2000US-0608408.
PR   03-AUG-2000; 2000US-0632366.
PR   21-SEP-2000; 2000US-0234687.
PR   27-SEP-2000; 2000US-0236359.
PR   04-OCT-2000; 2000GB-0024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for
PT   analyzing gene expression in human cervical epithelial cells -
XX
PS   Claim 27; SEQ ID No 20247; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP: see AAI10068-AAI28459). The present sequence is a peptide encoded
CC   by one such probe. The SENPs are derived from human HeLa cells. The SENPs
CC   can be used to produce a single exon microarray, which can be used for
CC   measuring human gene expression in a sample derived from human cervical
CC   epithelial cells. By measuring gene expression, the probes are therefore
CC   useful in grading and/or staging of diseases of the cervix, notably
CC   cervical cancer.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   24 AA;

  Query Match           44.2%;  Score 46;  DB 22;  Length 24;
  Best Local Similarity   63.6%;  Pred. No. 1.7;
  Matches    7;  Conservative    0;  Mismatches    4;  Indels    0;  Gaps    0;

Qy      7 HPCSXXGCRPG 17
          |||   || ||
Db      7 hpccgrgcwpg 17


Search completed: August 26, 2002, 13:36:50
Job time: 379 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:30:27 ; Search time 23.73 Seconds
(without alignments)
72.887 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	58.7	416	2	A32947	filaggrin precurso
2	61	58.7	2248	2	A35938	profilaggrin - hum
3	54	51.9	591	2	A45135	profilaggrin - hum
4	51	49.0	275	2	A36415	32K protein - vacc
5	51	49.0	328	2	S35336	transcription fact
6	51	49.0	377	2	T28558	hypothetical prote
7	51	49.0	377	2	H36849	A16L protein - var
8	51	49.0	377	2	T37403	35K myristylprotei
9	51	49.0	377	2	F72165	A17L protein - var
10	51	49.0	378	2	I42518	A16L protein - vac
11	50.5	48.6	207	2	S60006	Mad4 protein - mou
12	50.5	48.6	810	2	T10756	Nel-homolog protei
13	49.5	47.6	396	1	KXBOZ	plasma protein Z -
14	47.5	45.7	422	1	KXHUZ	plasma protein Z p
15	47	45.2	372	2	T45524	regulatory protein
16	47	45.2	2178	2	S29237	calcium channel pr
17	47	45.2	2222	2	A37490	voltage-dependent
18	47	45.2	2251	2	B54972	voltage-dependent
19	47	45.2	2259	2	S29236	calcium channel pr
20	47	45.2	2270	2	A54972	voltage-dependent
21	47	45.2	2272	2	C54972	voltage-dependent
22	46.5	44.7	2318	2	S45306	notch 3 protein -
23	46	44.2	221	2	T15845	hypothetical prote
24	46	44.2	397	2	H75066	GTP-binding protei
25	44.5	42.8	3051	2	S42373	hypothetical prote
26	44	42.3	391	2	A97633	hypothetical prote
27	44	42.3	391	2	AD2856	conserved hypothet
28	44	42.3	397	2	H71165	probable GTP-bindi
29	44	42.3	537	1	YRMSB6	tyrosinase-related
30	44	42.3	942	2	B72015	metalloproteinase,
31	44	42.3	942	2	C86610	insulinase family/
32	44	42.3	1224	2	E71611	hypothetical prote
33	43.5	41.8	2180	2	T29764	hypothetical prote
34	43.5	41.8	2437	2	S42612	transmembrane prot
35	43.5	41.8	2907	2	A57278	fibrillin-2 precur
36	43	41.3	59	2	T22272	hypothetical prote
37	42.5	40.9	473	2	A56175	adhesive plaque pr
38	42.5	40.9	485	2	JQ1957	glucagon receptor
39	42.5	40.9	685	2	JC7570	Delta-4 protein -
40	42.5	40.9	1296	2	T16859	hypothetical prote
41	42.5	40.9	2524	2	A35844	Xotch protein - Af
42	42.5	40.9	2531	2	S18188	notch protein homo
43	42.5	40.9	2531	2	A46019	Notch-1 protein -
44	42.5	40.9	2555	2	A40043	notch protein homo
45	42	40.4	237	2	S08073	cyclic nucleotide

ALIGNMENTS

RESULT 1
A32947
filaggrin precursor - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Dec-1989 #sequence_revision 04-Sep-1992 #text_change 29-Sep-1999
C;Accession: A32947
R;McKinley-Grant, L.J.; Idler, W.W.; Bernstein, I.A.; Parry, D.A.D.; Cannizzaro, L.;
Proc. Natl. Acad. Sci. U.S.A. 86, 4848-4852, 1989
A;Title: Characterization of a cDNA clone encoding human filaggrin and localization o
A;Reference number: A32947; MUID:89296901
A;Accession: A32947
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-416 <MCK>
A;Cross-references: GB:M24355; NID:g182604; PIDN:AAA52454.1; PID:g182605
A;Note: the authors translated the codon CAC for residue 188 as Gln, and AAT for resi
C;Genetics:
A;Gene: GDB:FLG
A;Cross-references: GDB:119912; OMIM:135940
A;Map position: 1q21-1q21
C;Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C;Keywords: EF hand; epidermis; polymorphism; tandem repeat

Query Match 58.7%; Score 61; DB 2; Length 416;
Best Local Similarity 66.7%; Pred. No. 0.1;
Matches 12; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```
Qy      1 QDTIHGHPCSXXGCRPGY 18
          |||| ||| |  | | ||
Db    395 QDTIRGHPGSSRGGRQGY 412
```

RESULT 2
A35938
profilaggrin - human (fragments)
C;Species: Homo sapiens (man)
C;Date: 14-Dec-1990 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
C;Accession: A35938
R;Gan, S.Q.; McBride, O.W.; Idler, W.W.; Markova, N.; Steinert, P.M.
Biochemistry 29, 9432-9440, 1990
A;Title: Organization, structure, and polymorphisms of the human profilaggrin gene.
A;Reference number: A35938; MUID:91064347
A;Accession: A35938
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-2248 <GAN>
A;Cross-references: GB:J02929
C;Genetics:
A;Gene: GDB:FLG
A;Cross-references: GDB:119912; OMIM:135940
A;Map position: 1q21-1q21

```
C;Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C;Keywords: EF hand; epidermis; polymorphism; tandem repeat
F;246-569/Region: filaggrin repeat
F;570-893/Region: filaggrin repeat
F;1074-1397/Region: filaggrin repeat
F;1573-1896/Region: filaggrin repeat


  Query Match             58.7%;  Score 61;  DB 2;  Length 2248;
  Best Local Similarity   66.7%;  Pred. No. 0.46;
  Matches   12;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPGY 18
          ||||| || |  | | ||
Db    291 QDTIHAHPGSRRGGRHGY 308


RESULT   3
A45135
profilaggrin - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 30-Apr-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C;Accession: A45135
R;Presland, R.B.; Haydock, P.V.; Fleckman, P.; Nirunsuksiri, W.; Dale, B.A.
J. Biol. Chem. 267, 23772-23781, 1992
A;Title: Characterization of the human epidermal profilaggrin gene. Genomic organizatio
A;Reference number: A45135; MUID:93054736
A;Accession: A45135
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: DNA
A;Residues: 1-591 <PRE>
A;Cross-references: GB:L01089; GB:M90967; NID:g190408; PIDN:AAA60177.1; PID:g553621
A;Note: sequence extracted from NCBI backbone (NCBIP:118773)
C;Genetics:
A;Gene: GDB:FLG
A;Cross-references: GDB:119912; OMIM:135940
A;Map position: 1q21-1q21
C;Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C;Keywords: EF hand; epidermis; polymorphism; tandem repeat
F;49-81/Domain: calmodulin repeat homology <EF2>


  Query Match             51.9%;  Score 54;  DB 2;  Length 591;
  Best Local Similarity   64.7%;  Pred. No. 1.6;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||| ||| |  | | |
Db    513 QDTIRGHPGSSRGGRQG 529


RESULT   4
A36415
32K protein - vaccinia virus (strain WR) (fragment)
C;Species: vaccinia virus
C;Date: 26-Jul-1991 #sequence_revision 26-Jul-1991 #text_change 21-Jul-2000
C;Accession: A36415
R;Pacha, R.F.; Meis, R.J.; Condit, R.C.
J. Virol. 64, 3853-3863, 1990
A;Title: Structure and expression of the vaccinia virus gene which prevents virus-induce
A;Reference number: A36415; MUID:90317884
A;Accession: A36415
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-275 <PAC>
A;Cross-references: EMBL:M32064; NID:g335834; PIDN:AAA48348.2; PID:g7555635


  Query Match             49.0%;  Score 51;  DB 2;  Length 275;
  Best Local Similarity   64.3%;  Pred. No. 2.3;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRPG 17
            ||| |||    |||
Db     85 IHGEPCSSFKFRPG 98


RESULT   5
S35336
transcription factor NF-M - chicken
N;Alternate names: C/EBP-beta protein homolog
C;Species: Gallus gallus (chicken)
C;Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 29-Sep-1999
C;Accession: S35336; S35321; S32116
R;Katz, S.; Kowenz-Leutz, E.; Mueller, C.; Meese, K.; Ness, S.A.; Leutz, A.
EMBO J. 12, 1321-1332, 1993
A;Title: The NF-M transcription factor is related to C/EBPbeta and plays a role in si
A;Reference number: S35336; MUID:93223673
A;Accession: S35336
A;Molecule type: mRNA
A;Residues: 1-328 <KAT>
A;Cross-references: EMBL:Z21646; NID:g296511; PIDN:CAA79760.1; PID:g296512
R;Burk, O.; Mink, S.; Ringwald, M.; Klempnauer, K.H.
EMBO J. 12, 2027-2038, 1993
A;Title: Synergistic activation of the chicken mim-1 gene by v-myb and C/EBP transcri
A;Reference number: S35321; MUID:93259145
A;Accession: S35321
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-328 <BUR>
A;Cross-references: EMBL:X70813; NID:g311999; PIDN:CAA50144.1; PID:g312000
C;Superfamily: CCAAT/enhancer-binding protein alpha
C;Keywords: DNA binding; leucine zipper; signal transduction; transcription regulatio


  Query Match             49.0%;  Score 51;  DB 2;  Length 328;
  Best Local Similarity   53.3%;  Pred. No. 2.7;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 TIHGHPCSXXGCRPG 17
          | |||||     :||
Db    128 TRHGHPCGSQSHKPG 142


RESULT   6
T28558
hypothetical protein A17L - variola major virus
C;Species: variola major virus
C;Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C;Accession: T28558
R;Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Au
Nature 366, 748-751, 1993
A;Title: Potential virulence determinants in terminal regions of variola smallpox vir
A;Reference number: Z20488; MUID:94088747
A;Accession: T28558
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-377 <MAS>
A;Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60868.1; PID:g439038
A;Experimental source: strain Bangladesh-1975


  Query Match             49.0%;  Score 51;  DB 2;  Length 377;
  Best Local Similarity   64.3%;  Pred. No. 3;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRPG 17
            ||| |||    |||
Db     85 IHGEPCSSFKFRPG 98


RESULT   7
H36849
A16L protein - variola virus (strain India-1967)
C;Species: variola virus
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 23-Mar-2001
```

C;Accession: H36849
R;Blinov, V.M.
submitted to GenBank, November 1992
A;Reference number: A36859
A;Accession: H36849
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-377 <BLI>
A;Cross-references: GB:X69198; NID:g456758; PIDN:CAA49061.1; PID:g297299

Query Match 49.0%; Score 51; DB 2; Length 377;
Best Local Similarity 64.3%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 IHGHPCSXXGCRPG 17
||| ||| ||| |||
Db 85 IHGEPCSSFKFRPG 98

RESULT 8
T37403
35K myristylprotein - vaccinia virus (strain Ankara)
C;Species: vaccinia virus
A;Variety: strain Ankara
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T37403
R;Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.
submitted to the EMBL Data Library, March 1997
A;Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strai
A;Reference number: Z20877
A;Accession: T37403
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-377 <ANT>
A;Cross-references: EMBL:U94848; PIDN:AAB96467.1
A;Experimental source: strain Ankara
C;Genetics:
A;Note: MVA127L

Query Match 49.0%; Score 51; DB 2; Length 377;
Best Local Similarity 64.3%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 IHGHPCSXXGCRPG 17
||| ||| ||| |||
Db 85 IHGEPCSSFKFRPG 98

RESULT 9
F72165
A17L protein - variola minor virus (strain Garcia-1966)
C;Species: variola minor virus
C;Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
C;Accession: F72165
R;Shchelkunov, S.N.; Totmenin, A.V.; Gutorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar
submitted to GenBank, March 1998
A;Description: Analysis of the complete coding sequence of DNA of alastrim variola minor
A;Reference number: A72150
A;Accession: F72165
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-377 <SHC>
A;Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54720.1; PID:g5830681
A;Experimental source: strain Garcia-1966
C;Genetics:
A;Gene: A17L

Query Match 49.0%; Score 51; DB 2; Length 377;
Best Local Similarity 64.3%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 IHGHPCSXXGCRPG 17
||| ||| ||| |||
Db 85 IHGEPCSSFKFRPG 98

RESULT 10
I42518
A16L protein - vaccinia virus (strain Copenhagen)
C;Species: vaccinia virus
A;Note: host Homo sapiens (man)
C;Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 08-Apr-1994
C;Accession: I42518
R;Johnson, G.P.
submitted to GenBank, June 1990
A;Reference number: A33172
A;Accession: I42518
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <JOH>

Query Match 49.0%; Score 51; DB 2; Length 378;
Best Local Similarity 64.3%; Pred. No. 3;
Matches 9; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 4 IHGHPCSXXGCRPG 17
||| ||| ||| |||
Db 85 IHGEPCSSFKFRPG 98

RESULT 11
S60006
Mad4 protein - mouse
C;Species: Mus musculus (house mouse)
C;Date: 23-Aug-1996 #sequence_revision 13-Mar-1997 #text_change 18-Aug-2000
C;Accession: S60006
R;Hurlin, P.J.; Queva, C.; Koskinen, P.J.; Steingrimsson, E.; Ayer, D.E.; Copeland, N
EMBO J. 14, 5646-5659, 1995
A;Title: Mad3 and Mad4: novel Max-interacting transcriptional repressors that suppres
A;Reference number: S60005; MUID:96091137
A;Accession: S60006
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-207 <HUR>
C;Superfamily: human Max-interacting protein 1

Query Match 48.6%; Score 50.5; DB 2; Length 207;
Best Local Similarity 56.2%; Pred. No. 2.1;
Matches 9; Conservative 2; Mismatches 4; Indels 1; Gaps 1;

Qy 2 DTIHGHPCSXXGCRPG 17
|: :|||| || ||
Db 191 DSSYGHPCRRPGC-PG 205

RESULT 12
T10756
Nel-homolog protein - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999
C;Accession: T10756
R;Kuroda, S.; Tokunaga, C.; Kiyohara, Y.; Konishi, H.; Matsuhashi, S.; Kikkawa, U.
submitted to the EMBL Data Library, November 1998
A;Description: Protein kinase C-binding protein.
A;Reference number: Z17122
A;Accession: T10756
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-810 <KUR>
A;Cross-references: EMBL:U48246; NID:g3851179; PID:g3851180
A;Experimental source: strain Sprague-Dawley, brain

```
  Query Match             48.6%;  Score 50.5;  DB 2;  Length 810;
  Best Local Similarity   47.1%;  Pred. No. 7.1;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    3;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPGY 18
           :|: || |:   |:|||
Db     495 NTVQGHSCT---CQPGY 508


RESULT  13
KXBOZ
plasma protein Z - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 16-Jul-1999
C;Accession: A22171; A00926
R;Hoejrup, P.; Jensen, M.S.; Petersen, T.E.
FEBS Lett. 184, 333-338, 1985
A;Title: Amino acid sequence of bovine protein Z: a vitamin K-dependent serine protease
A;Reference number: A22171; MUID:85204370
A;Accession: A22171
A;Molecule type: protein
A;Residues: 1-396 <HOE>
C;Comment: Protein Z is a single-chain plasma glycoprotein of unknown function. Although
and has no enzymatic activity.
C;Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology
C;Keywords: beta-hydroxyaspartic acid; calcium binding; carboxyglutamic acid; glycoprote
F;1-46/Domain: calcium binding #status predicted <CAB>
F;1-45/Domain: Gla domain homology (fragment) <GLA>
F;51-82/Domain: EGF homology <EG1>
F;89-125/Domain: EGF homology <EG2>
F;143-352/Domain: trypsin homology <TRY>
F;7,8,11,15,17,20,21,26,27,30,33,36,40/Modified site: gamma-carboxyglutamic acid (Glu) #
F;59,191,289/Binding site: carbohydrate (Asn) (covalent) #status experimental
F;64/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F;388/Binding site: carbohydrate (Thr) (covalent) #status experimental


  Query Match             47.6%;  Score 49.5;  DB 1;  Length 396;
  Best Local Similarity   50.0%;  Pred. No. 5.3;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||:| |: |:   | |||
Db      63 QDSIRGYACT---CAPGY 77


RESULT  14
KXHUZ
plasma protein Z precursor [validated] - human
N;Alternate names: vitamin K-dependent glycoprotein Z
C;Species: Homo sapiens (man)
C;Date: 18-Jan-1991 #sequence_revision 05-Jan-1996 #text_change 08-Dec-2000
C;Accession: A36244; A35893; B35893
R;Ichinose, A.; Takeya, H.; Espling, E.; Iwanaga, S.; Kisiel, W.; Davie, E.W.
Biochem. Biophys. Res. Commun. 172, 1139-1144, 1990
A;Title: Amino acid sequence of human protein Z, a vitamin K-dependent plasma glycoprote
A;Reference number: A36244; MUID:91058548
A;Accession: A36244
A;Molecule type: mRNA
A;Residues: 1-422 <ICH>
A;Cross-references: GB:M55671; NID:g190465; PIDN:AAA36501.1; PID:g190466
A;Note: parts of this sequence, including the amino end of the mature protein, were dete
R;Sejima, H.; Hayashi, T.; Deyashiki, Y.; Nishioka, J.; Suzuki, K.
Biochem. Biophys. Res. Commun. 171, 661-668, 1990
A;Title: Primary structure of vitamin K-dependent human protein Z.
A;Reference number: A35893; MUID:90386637
A;Accession: A35893
A;Molecule type: protein
A;Residues: 63-68,'XX',71-72,'X',74-76,'X',78,'XX',81,'XX',84,'X',86-87,'XX',90,'XX',93-
A;Accession: B35893
A;Molecule type: mRNA
A;Residues: 103-422 <SE2>
A;Cross-references: GB:M59303; NID:g190461; PIDN:AAA36499.1; PID:g190462
C;Genetics:
A;Gene: GDB:PROZ
A;Cross-references: GDB:9957440; OMIM:176895
A;Map position: 13q34-13q34
C;Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol
C;Keywords: beta-hydroxyaspartic acid; calcium binding; carboxyglutamic acid; glycopr
F;1-10/Domain: signal sequence #status predicted <SIG>
F;11-62/Domain: propeptide #status predicted <PRO>
F;47-107/Domain: Gla domain homology <GLA>
F;63-422/Product: protein Z #status experimental <MAT>
F;113-144/Domain: EGF homology <EG1>
F;151-187/Domain: EGF homology <EG2>
F;199-417/Domain: trypsin homology <TRY>
F;69,70,73,77,79,82,83,88,89,92,95,97,102/Modified site: gamma-carboxyglutamic acid (
F;115/Binding site: carbohydrate (Ser) (covalent) #status experimental
F;121,247,255,328,354/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;126/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted
F;134,337/Binding site: carbohydrate (Thr) (covalent) #status predicted
F;225-241,349-363/Disulfide bonds: #status predicted
F;258/Binding site: carbohydrate (Ser) (covalent) #status predicted


  Query Match             45.7%;  Score 47.5;  DB 1;  Length 422;
  Best Local Similarity   50.0%;  Pred. No. 11;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||:| |: |:   | |||
Db     125 QDSIWGYTCT---CSPGY 139


RESULT  15
T45524
regulatory protein rim101 homolog [imported] - yeast (Kluyveromyces marxianus var. la
C;Species: Kluyveromyces marxianus var. lactis, Candida sphaerica
C;Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 31-Jan-2000
C;Accession: T45524
R;Bao, W.G.; Fukuhara, H.
submitted to the EMBL Data Library, July 1999
A;Description: The ubiquitin-encoding genes of Kluyveromyces lactis.
A;Reference number: Z23000
A;Accession: T45524
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-372 <BAO>
A;Cross-references: EMBL:AJ243800; PIDN:CAB50896.1
A;Experimental source: strain 2359/152
C;Genetics:
A;Gene: rim101


  Query Match             45.2%;  Score 47;  DB 2;  Length 372;
  Best Local Similarity   50.0%;  Pred. No. 12;
  Matches    7;  Conservative    2;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGC 14
           :|||:| | |   ||
Db     105 EDTVHLHRCQWKGC 118


Search completed: August 26, 2002, 13:30:27
Job time: 152 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:35:49 ; Search time 17.88 Seconds
(without alignments)
38.979 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	58.7	416	1	FILA_HUMAN	P20930 homo sapien
2	51.5	49.5	810	1	NEL1_HUMAN	Q92832 homo sapien
3	51	49.0	275	1	VA16_VACCV	P16710 vaccinia vi
4	51	49.0	328	1	CEBB_CHICK	Q05826 gallus gall
5	51	49.0	377	1	VA16_VARV	P33841 variola vir
6	51	49.0	378	1	VA16_VACCC	P20993 vaccinia vi
7	50.5	48.6	209	1	MAD4_MOUSE	Q60948 mus musculu
8	50.5	48.6	810	1	NEL1_RAT	Q62919 rattus norv
9	49.5	47.6	396	1	PRTZ_BOVIN	P00744 bos taurus
10	47.5	45.7	400	1	PRTZ_HUMAN	P22891 homo sapien
11	47	45.2	2222	1	CCAE_RAT	Q07652 rattus norv
12	47	45.2	2259	1	CCAE_RABIT	Q02343 oryctolagus
13	47	45.2	2272	1	CCAE_MOUSE	Q61290 mus musculu
14	47	45.2	2312	1	CCAE_HUMAN	Q15878 homo sapien
15	46.5	44.7	2318	1	NTC3_MOUSE	Q61982 mus musculu
16	45.5	43.8	816	1	NEL2_HUMAN	Q99435 homo sapien
17	45.5	43.8	816	1	NEL2_MOUSE	Q61220 mus musculu
18	45.5	43.8	816	1	NEL2_RAT	Q62918 rattus norv
19	44.5	42.8	3051	1	YNX3_CAEEL	P34576 caenorhabdi
20	44	42.3	537	1	TYR1_MOUSE	P07147 mus musculu
21	44	42.3	830	1	SREC_HUMAN	Q14162 homo sapien
22	43.5	41.8	2437	1	NOTC_BRARE	P46530 brachydanio
23	43.5	41.8	2907	1	FBN2_MOUSE	Q61555 mus musculu
24	43	41.3	272	1	Y4PM_RHISN	P55618 rhizobium s
25	42.5	40.9	368	1	LNK_RAT	P50745 rattus norv
26	42.5	40.9	473	1	FP2_MYTGA	Q25464 mytilus gal
27	42.5	40.9	485	1	GLR_RAT	P30082 rattus norv
28	42.5	40.9	685	1	DLL4_HUMAN	Q9nr61 homo sapien
29	42.5	40.9	2444	1	NTC1_HUMAN	P46531 homo sapien
30	42.5	40.9	2524	1	NOTC_XENLA	P21783 xenopus lae
31	42.5	40.9	2531	1	NTC1_MOUSE	Q01705 mus musculu
32	42.5	40.9	2531	1	NTC1_RAT	Q07008 rattus norv
33	42.5	40.9	4655	1	LRP2_HUMAN	P98164 homo sapien
34	42	40.4	237	1	IPDE_DICDI	P22549 dictyosteli
35	42	40.4	2569	1	LMA3_MOUSE	Q61789 mus musculu
36	41.5	39.9	492	1	FA10_BOVIN	P00743 bos taurus
37	41.5	39.9	1064	1	FBP1_STRPU	P10079 strongyloce
38	41.5	39.9	4660	1	LRP2_RAT	P98158 rattus norv
39	41	39.4	130	1	OREX_MOUSE	O55241 mus musculu
40	41	39.4	250	1	YK21_YEAST	P36134 saccharomyc
41	41	39.4	1106	1	GLI1_HUMAN	P08151 homo sapien
42	41	39.4	1111	1	GLI1_MOUSE	P47806 mus musculu
43	41	39.4	1226	1	POLG_DEN1W	P17763 dengue viru
44	41	39.4	1393	1	RPOC_CHLPN	Q9z999 chlamydia p
45	41	39.4	1396	1	RPOC_CHLMU	Q9pk79 chlamydia m

ALIGNMENTS

RESULT 1
FILA_HUMAN
ID FILA_HUMAN STANDARD; PRT; 416 AA.
AC P20930;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Filaggrin precursor (Fragment).
GN FLG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89296901; PubMed=2740331;
RA McKinley-Grant L.J., Idler W.W., Bernstein I.A., Parry D.A.D.,
RA Cannizzaro L., Croce C.M., Huebner K., Lessin S.R., Steinert P.M.;
RT "Characterization of a cDNA clone encoding human filaggrin and
RT localization of the gene to chromosome region 1q21.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:4848-4852(1989).
RN [2]
RP CITRULLINATION.
RX MEDLINE=96374388; PubMed=8780679;
RA Senshu T., Kan S., Ogawa H., Manabe M., Asaga H.;
RT "Preferential deimination of keratin K1 and filaggrin during the
RT terminal differentiation of human epidermis.";
RL Biochem. Biophys. Res. Commun. 225:712-719(1996).
CC -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC -!- PTM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC OF 324 AA, WHICH ARE NOT SEPARATED BY "LARGE LINKER". THE
CC PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC -!- PTM: Undergoes deimination of some arginine residues
CC (citrullination).
CC ----------

CC ----------
DR EMBL; M24355; AAA52454.1; -.
DR PIR; A32947; A32947.
DR MIM; 135940; -.
DR InterPro; IPR003303; Filaggrin.
DR PRINTS; PR00487; FILAGGRIN.
KW Phosphorylation; Citrullination; Developmental protein.
FT NON_TER 1 1
SQ SEQUENCE 416 AA; 44105 MW; DEEA3218BA043F32 CRC64;

```
  Query Match             58.7%;  Score 61;  DB 1;  Length 416;
  Best Local Similarity   66.7%;  Pred. No. 0.018;
  Matches   12;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           |||| ||| |   | | ||
Db     395 QDTIRGHPGSSRGGRQGY 412


RESULT   2
NEL1_HUMAN
ID   NEL1_HUMAN     STANDARD;      PRT;   810 AA.
AC   Q92832; Q9Y472;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein kinase C-binding protein NELL1 precursor (NEL-like protein 1)
DE   (Nel-related protein 1).
GN   NELL1 OR NRP1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=97131504; PubMed=8975702;
RA   Watanabe T.K., Katagiri T., Suzuki M., Shimizu F., Fujiwara T.,
RA   Kanemoto N., Nakamura Y., Hirai Y., Maekawa H., Takahashi E.;
RT   "Cloning and characterization of two novel human cDNAs (NELL1 and
RT   NELL2) encoding proteins with six EGF-like repeats.";
RL   Genomics 38:273-276(1996).
RN   [2]
RP   SEQUENCE OF 383-810 FROM N.A.
RA   Ting K., Vastardis H., Mulliken J.B., Bertolami C., Wen Z.,
RA   Young M., Tieu A., Kwong E.;
RT   "Nel homolog gene expression in craniofacial anomalies.";
RL   Submitted (SEP-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1 (BY SIMILARITY).
CC   -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC   -!- DISEASE: EXPRESSED IN CRANIOFACIAL ANOMALIES.
CC   -!- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
CC       IN POSITIONS 427 AND 771.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; D83017; BAA11680.1; -.
DR   EMBL; U57523; AAB06946.1; ALT_FRAME.
DR   HSSP; P07204; 1ADX.
DR   MIM; 602319; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF00093; vwc; 2.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00001; EGF_like; 4.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 3.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   810       PROTEIN KINASE C-BINDING PROTEIN NELL1.
FT   DOMAIN      81   230       TSP N-TERMINAL.
FT   DOMAIN     273   331       VWFC 1.
FT   DOMAIN     335   390       VWFC 2.
FT   DOMAIN     391   433       EGF-LIKE 1.
FT   DOMAIN     434   475       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     476   516       EGF-LIKE 3.
FT   DOMAIN     515   547       EGF-LIKE 4.
FT   DOMAIN     549   595       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     596   631       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632   687       VWFC 3.
FT   DOMAIN     692   750       VWFC 4.
FT   DOMAIN     752   807       VWFC 5.
FT   DISULFID   395   407       BY SIMILARITY.
FT   DISULFID   401   416       BY SIMILARITY.
FT   DISULFID   418   432       BY SIMILARITY.
FT   DISULFID   438   451       BY SIMILARITY.
FT   DISULFID   445   460       BY SIMILARITY.
FT   DISULFID   462   474       BY SIMILARITY.
FT   DISULFID   480   493       BY SIMILARITY.
FT   DISULFID   487   502       BY SIMILARITY.
FT   DISULFID   504   515       BY SIMILARITY.
FT   DISULFID   519   529       BY SIMILARITY.
FT   DISULFID   523   535       BY SIMILARITY.
FT   DISULFID   537   546       BY SIMILARITY.
FT   DISULFID   553   566       BY SIMILARITY.
FT   DISULFID   560   575       BY SIMILARITY.
FT   DISULFID   577   594       BY SIMILARITY.
FT   DISULFID   600   613       BY SIMILARITY.
FT   DISULFID   607   622       BY SIMILARITY.
FT   DISULFID   624   630       BY SIMILARITY.
FT   CARBOHYD    40    40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    53    53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    83    83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   224   224       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   294   294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   372   372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   511   511       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   562   562       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   609   609       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   708   708       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   732   732       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   758   758       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CONFLICT   383   383       N -> D (IN REF. 2).
FT   CONFLICT   573   573       Y -> H (IN REF. 2).
FT   CONFLICT   626   626       S -> C (IN REF. 2).
SQ   SEQUENCE   810 AA;  89606 MW;  549465EA3F7AEED0 CRC64;


  Query Match             49.5%;  Score 51.5;  DB 1;  Length 810;
  Best Local Similarity   47.1%;  Pred. No. 1;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    3;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPGY 18
           :|: || |:   |:|||
Db     495 NTVQGHSCT---CKPGY 508


RESULT   3
VA16_VACCV
ID   VA16_VACCV     STANDARD;      PRT;   275 AA.
AC   P16710;
DT   01-AUG-1990 (Rel. 15, Created)
DT   01-AUG-1990 (Rel. 15, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A16 (Fragment).
```

```
GN   A16L.
OS   Vaccinia virus (strain WR).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10254;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90317884; PubMed=2370683;
RA   Pacha R.F., Meis R.J., Condit R.C.;
RT   "Structure and expression of the vaccinia virus gene which prevents
RT   virus-induced breakdown of RNA.";
RL   J. Virol. 64:3853-3863(1990).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A16 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M32064; AAA48348.2; -.
DR   PIR; A36415; A36415.
FT   NON_TER     275    275
SQ   SEQUENCE   275 AA;  31811 MW;  E2461AB1DB7B93A3 CRC64;


  Query Match            49.0%;  Score 51;  DB 1;  Length 275;
  Best Local Similarity  64.3%;  Pred. No. 0.45;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRPG 17
          ||| |||    |||
Db     85 IHGEPCSSFKFRPG 98


RESULT   4
CEBB_CHICK
ID   CEBB_CHICK     STANDARD;      PRT;   328 AA.
AC   Q05826;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   01-FEB-1995 (Rel. 31, Last annotation update)
DE   CCAAT/enhancer binding protein beta (C/EBP beta) (Transcription
DE   factor NF-M) (CCR protein).
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93223673; PubMed=8467792;
RA   Katz S., Kowenz-Leutz E., Mueller C., Meese K., Ness S.A.,
RA   Leutz A.;
RT   "The NF-M transcription factor is related to C/EBP beta and plays a
RT   role in signal transduction, differentiation and leukemogenesis of
RT   avian myelomonocytic cells.";
RL   EMBO J. 12:1321-1332(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93259145; PubMed=8491193;
RA   Burk O., Mink S., Ringwald M., Klempnauer K.H.;
RT   "Synergistic activation of the chicken mim-1 gene by v-myb and C/EBP
RT   transcription factors.";
RL   EMBO J. 12:2027-2038(1993).
CC   -!- FUNCTION: HAS A ROLE IN SIGNAL TRANSDUCTION, DIFFERENTIATION AND
CC       LEUKEMOGENESIS OF MYELOMONOCYTIC CELLS. BINDS TO THE MGF AND MIM-1
CC       PROMOTERS AND ACTIVATES THE TRANSCRIPTION OF THESE GENES.
CC   -!- SUBUNIT: BINDS DNA AS A DIMER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN MYELOMONCYTIC
CC       CELLS.
CC   -!- SIMILARITY: TO OTHER BZIP PROTEINS. STRONG, TO OTHER C/EBP
CC       PROTEINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; Z21646; CAA79760.1; -.
DR   EMBL; X70813; CAA50144.1; -.
DR   PIR; S35336; S35336.
DR   TRANSFAC; T02022; -.
DR   InterPro; IPR001871; bZIP.
DR   Pfam; PF00170; bZIP; 1.
DR   SMART; SM00338; BRLZ; 1.
KW   Transcription regulation; Activator; DNA-binding; Nuclear protein.
FT   DNA_BIND     260    276       BASIC MOTIF.
FT   DOMAIN       289    317       LEUCINE-ZIPPER.
SQ   SEQUENCE   328 AA;  35030 MW;  5AAE257F8213671C CRC64;


  Query Match            49.0%;  Score 51;  DB 1;  Length 328;
  Best Local Similarity  53.3%;  Pred. No. 0.53;
  Matches    8;  Conservative    1;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      3 TIHGHPCSXXGCRPG 17
          | |||||     :||
Db    128 TRHGHPCGSQSHKPG 142


RESULT   5
VA16_VARV
ID   VA16_VARV      STANDARD;      PRT;   377 AA.
AC   P33841;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A16.
GN   A16L.
OS   Variola virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=INDIA-1967 / ISOLATE IND3;
RX   MEDLINE=93202281; PubMed=8384129;
RA   Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT   "Genes of variola and vaccinia viruses necessary to overcome the host
RT   protective mechanisms.";
RL   FEBS Lett. 319:80-83(1993).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A16 FAMILY.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; X69198; CAA49061.1; -.
DR   PIR; H36849; H36849.
DR   InterPro; IPR004251; DUF230.
DR   Pfam; PF03003; DUF230; 1.
SQ   SEQUENCE   377 AA;  43545 MW;  98002E9E45602894 CRC64;


  Query Match            49.0%;  Score 51;  DB 1;  Length 377;
```

```
  Best Local Similarity   64.3%;  Pred. No. 0.61;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 IHGHPCSXXGCRPG 17
           ||| |||    |||
Db      85 IHGEPCSSFKFRPG 98


RESULT   6
VA16_VACCC
ID   VA16_VACCC     STANDARD;      PRT;   378 AA.
AC   P20993;
DT   01-FEB-1991 (Rel. 17, Created)
DT   01-FEB-1991 (Rel. 17, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein A16.
GN   A16L.
OS   Vaccinia virus (strain Copenhagen).
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10249;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91021027; PubMed=2219722;
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "The complete DNA sequence of vaccinia virus.";
RL   Virology 179:247-266(1990).
RN   [2]
RP   COMPLETE GENOME.
RA   Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA   Paoletti E.;
RT   "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL   Virology 179:517-563(1990).
CC   -!- SIMILARITY: BELONGS TO THE POXVIRUSES A16 FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M35027; AAA48138.1; -.
DR   PIR; I42518; I42518.
DR   InterPro; IPR004251; DUF230.
DR   Pfam; PF03003; DUF230; 1.
SQ   SEQUENCE   378 AA;  43561 MW;  05ED614AA1D11A19 CRC64;


  Query Match             49.0%;  Score 51;  DB 1;  Length 378;
  Best Local Similarity   64.3%;  Pred. No. 0.61;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 IHGHPCSXXGCRPG 17
           ||| |||    |||
Db      85 IHGEPCSSFKFRPG 98


RESULT   7
MAD4_MOUSE
ID   MAD4_MOUSE     STANDARD;      PRT;   209 AA.
AC   Q60948;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-JUL-1999 (Rel. 38, Last annotation update)
DE   MAX-interacting transcriptional repressor MAD4.
GN   MAD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB1;
RX   MEDLINE=96091137; PubMed=8521822;
RA   Hurlin P.J., Queva C., Koskinen P.J., Steingrimsson E., Ayer D.E.,
RA   Copeland N.G., Jenkins N.A., Eisenman R.N.;
RT   "Mad3 and Mad4: novel Max-interacting transcriptional repressors that
RT   suppress c-myc dependent transformation and are expressed during
RT   neural and epidermal differentiation.";
RL   EMBO J. 14:5646-5659(1995).
CC   -!- FUNCTION: TRANSCRIPTIONAL REPRESSOR. MAD4 BINDS WITH MAX TO FORM A
CC       SEQUENCE-SPECIFIC DNA-BINDING PROTEIN COMPLEX WHICH RECOGNIZES THE
CC       CORE SEQUENCE CAC[GA]TG. MAD4 THUS ANTAGONIZES MYC TRANSCRIPTIONAL
CC       ACTIVITY BY COMPETING FOR MAX.
CC   -!- SUBUNIT: EFFICIENT DNA BINDING REQUIRES DIMERIZATION WITH ANOTHER
CC       BHLH PROTEIN. BINDS DNA AS A HETERODIMER WITH MAX.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
CC       TRANSCRIPTION FACTORS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U32395; AAB02795.1; -.
DR   TRANSFAC; T02392; -.
DR   MGD; MGI:104991; Mad4.
DR   InterPro; IPR003015; HLH_Myc.
DR   InterPro; IPR001092; HLH_dim.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
DR   PROSITE; PS00038; HELIX_LOOP_HELIX; FALSE_NEG.
DR   PROSITE; PS50888; Transcription regulation; Repressor.
KW   Nuclear protein; DNA-binding; Transcription regulation; Repressor.
FT   DNA_BIND      55     66       BASIC DOMAIN.
FT   DOMAIN        67    106       HELIX-LOOP-HELIX MOTIF (POTENTIAL).
SQ   SEQUENCE   209 AA;  23590 MW;  C8FB5A56AFE3B27C CRC64;


  Query Match             48.6%;  Score 50.5;  DB 1;  Length 209;
  Best Local Similarity   56.2%;  Pred. No. 0.42;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPG 17
           |: :|||   || ||
Db     193 DSSYGHPCRRPGC-PG 207


RESULT   8
NEL1_RAT
ID   NEL1_RAT       STANDARD;      PRT;   810 AA.
AC   Q62919;
DT   01-NOV-1997 (Rel. 35, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Protein kinase C-binding protein NELL1 precursor (NEL-like protein 1).
GN   NELL1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=20017976; PubMed=10548494;
RA   Kuroda S., Oyasu M., Kawakami M., Kanayama N., Tanizawa K., Saito N.,
RA   Abe T., Matsuhashi S., Ting K.;
RT   "Biochemical characterization and expression analysis of neural
RT   thrombospondin-1-like proteins NELL1 and NELL2.";
RL   Biochem. Biophys. Res. Commun. 265:79-86(1999).
```

```
CC   -!- SUBUNIT: HOMOTRIMER. BINDS TO PKC BETA-1.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- SIMILARITY: CONTAINS 1 TSP N-TERMINAL DOMAIN.
CC   -!- SIMILARITY: CONTAINS 5 VWFC DOMAINS.
CC   -!- SIMILARITY: CONTAINS 6 EGF-LIKE DOMAINS.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U48246; AAC72252.1; -.
DR   HSSP; P07204; 1ADX.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001791; Laminin_G.
DR   InterPro; IPR003129; TSPN.
DR   InterPro; IPR001007; VWFC.
DR   Pfam; PF00008; EGF; 4.
DR   Pfam; PF02210; TSPN; 1.
DR   Pfam; PF00093; vwc; 3.
DR   SMART; SM00179; EGF_CA; 2.
DR   SMART; SM00001; EGF_like; 4.
DR   SMART; SM00282; LamG; 1.
DR   SMART; SM00210; TSPN; 1.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 3.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 3.
DR   PROSITE; PS01187; EGF_CA; 3.
DR   PROSITE; PS01208; VWFC; 2.
KW   Glycoprotein; EGF-like domain; Repeat; Signal.
FT   SIGNAL       1    16       POTENTIAL.
FT   CHAIN       17   810       PROTEIN KINASE C-BINDING PROTEIN NELL1.
FT   DOMAIN      81   230       TSP N-TERMINAL.
FT   DOMAIN     273   331       VWFC 1.
FT   DOMAIN     335   390       VWFC 2.
FT   DOMAIN     391   433       EGF-LIKE 1.
FT   DOMAIN     434   475       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     476   516       EGF-LIKE 3.
FT   DOMAIN     515   547       EGF-LIKE 4.
FT   DOMAIN     549   595       EGF-LIKE 5, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     596   631       EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     632   687       VWFC 3.
FT   DOMAIN     692   750       VWFC 4.
FT   DOMAIN     752   807       VWFC 5.
FT   DISULFID   395   407       BY SIMILARITY.
FT   DISULFID   401   416       BY SIMILARITY.
FT   DISULFID   418   432       BY SIMILARITY.
FT   DISULFID   438   451       BY SIMILARITY.
FT   DISULFID   445   460       BY SIMILARITY.
FT   DISULFID   462   474       BY SIMILARITY.
FT   DISULFID   480   493       BY SIMILARITY.
FT   DISULFID   487   502       BY SIMILARITY.
FT   DISULFID   504   515       BY SIMILARITY.
FT   DISULFID   519   529       BY SIMILARITY.
FT   DISULFID   523   535       BY SIMILARITY.
FT   DISULFID   537   546       BY SIMILARITY.
FT   DISULFID   553   566       BY SIMILARITY.
FT   DISULFID   560   575       BY SIMILARITY.
FT   DISULFID   577   594       BY SIMILARITY.
FT   DISULFID   600   613       BY SIMILARITY.
FT   DISULFID   607   622       BY SIMILARITY.
FT   DISULFID   624   630       BY SIMILARITY.
FT   CARBOHYD    40    40       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    53    53       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    83    83       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   224   224       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   294   294       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   372   372       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   511   511       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   562   562       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   609   609       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   708   708       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   810 AA;  89212 MW;  46F09C466AF9AB0B CRC64;


  Query Match             48.6%;  Score 50.5;  DB 1;  Length 810;
  Best Local Similarity   47.1%;  Pred. No. 1.5;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    3;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPGY 18
           :|: || |:  |:|||
Db     495 NTVQGHSCT---CQPGY 508


RESULT   9
PRTZ_BOVIN
ID   PRTZ_BOVIN     STANDARD;      PRT;   396 AA.
AC   P00744;
DT   21-JUL-1986 (Rel. 01, Created)
DT   21-JUL-1986 (Rel. 01, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Vitamin K-dependent protein Z.
GN   PROZ.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE.
RX   MEDLINE=85204370; PubMed=3888670;
RA   Hoejrup P., Jensen M.S., Petersen T.E.;
RT   "Amino acid sequence of bovine protein Z: a vitamin K-dependent
RT   serine protease homolog.";
RL   FEBS Lett. 184:333-338(1985).
RN   [2]
RP   STRUCTURE OF CARBOHYDRATE ON SER-53.
RX   MEDLINE=90062160; PubMed=2511201;
RA   Nishimura H., Kawabata S., Kisiel W., Hase S., Ikenaka T., Takao T.,
RA   Shimonishi Y., Iwanaga S.;
RT   "Identification of a disaccharide (Xyl-Glc) and a trisaccharide
RT   (Xyl2-Glc) O-glycosidically linked to a serine residue in the first
RT   epidermal growth factor-like domain of human factors VII and IX and
RT   protein Z and bovine protein Z.";
RL   J. Biol. Chem. 264:20320-20325(1989).
RN   [3]
RP   STRUCTURE OF CARBOHYDRATE ON SER-53.
RX   MEDLINE=91344709; PubMed=2129367;
RA   Iwanaga S., Nishimura H., Kawabata S., Kisiel W., Hase S., Ikenaka T.;
RT   "A new trisaccharide sugar chain linked to a serine residue in the
RT   first EGF-like domain of clotting factors VII and IX and protein Z.";
RL   Adv. Exp. Med. Biol. 281:121-131(1990).
CC   -!- FUNCTION: APPEARS TO ASSIST HEMOSTASIS BY BINDING THROMBIN AND
CC       PROMOTING ITS ASSOCIATION WITH PHOSPHOLIPID VESICLES.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: ALTHOUGH HOMOLOGOUS WITH THE VITAMIN K-DEPENDENT
CC       CLOTTING FACTORS, IT HAS LOST TWO OF THE ESSENTIAL CATALYTIC
CC       RESIDUES AND HAS NO ENZYMATIC ACTIVITY.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
DR   PIR; A22171; KXBOZ.
DR   HSSP; P00740; 1CFH.
DR   MEROPS; S01.979; -.
DR   GlycoSuiteDB; P00744; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR001254; Trypsin.
```

```
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00001; GLABLOOD.
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
KW   Plasma; Glycoprotein; Gamma-carboxyglutamic acid; Hydroxylation;
KW   Calcium; Serine protease homolog; Vitamin K; EGF-like domain.
FT   DOMAIN       47     83       EGF-LIKE 1.
FT   DOMAIN       85    126       EGF-LIKE 2.
FT   DOMAIN      135    357       SERINE PROTEASE.
FT   MOD_RES       7      7       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES       8      8       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      11     11       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      15     15       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      17     17       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      20     20       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      21     21       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      26     26       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      27     27       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      30     30       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      33     33       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      36     36       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      40     40       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES      64     64       HYDROXYLATION.
FT   DISULFID     51     62       BY SIMILARITY.
FT   DISULFID     56     71       BY SIMILARITY.
FT   DISULFID     73     82       BY SIMILARITY.
FT   DISULFID     89    101       BY SIMILARITY.
FT   DISULFID     97    110       BY SIMILARITY.
FT   DISULFID    112    125       BY SIMILARITY.
FT   CARBOHYD     53     53       O-LINKED (GLC. . .).
FT                                /FTId=CAR_000032.
FT   CARBOHYD     59     59       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    191    191       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    289    289       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    388    388       O-LINKED.
SQ   SEQUENCE   396 AA;  43112 MW;  04C5D7A35849B116 CRC64;


  Query Match             47.6%;  Score 49.5;  DB 1;  Length 396;
  Best Local Similarity   50.0%;  Pred. No. 1.1;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||:| |: |:   | |||
Db      63 QDSIRGYACT---CAPGY 77


RESULT  10
PRTZ_HUMAN
ID   PRTZ_HUMAN     STANDARD;      PRT;   400 AA.
AC   P22891; Q15213;
DT   01-AUG-1991 (Rel. 19, Created)
DT   01-NOV-1991 (Rel. 20, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Vitamin K-dependent protein Z precursor.
GN   PROZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Liver;
RX   MEDLINE=91058548; PubMed=2244898;
RA   Ichinose A., Takeya H., Espling E., Iwanaga S., Kisiel W., Davie E.W.;
RT   "Amino acid sequence of human protein Z, a vitamin K-dependent plasma
RT   glycoprotein.";
RL   Biochem. Biophys. Res. Commun. 172:1139-1144(1990).
RN   [2]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=98244864; PubMed=9578570;
RA   Fujimaki K., Yamazaki T., Taniwaki M., Ichinose A.;
RT   "The gene for human protein Z is localized to chromosome 13 at band
RT   q34 and is coded by eight regular exons and one alternative exon.";
RL   Biochemistry 37:6838-6846(1998).
RN   [3]
RP   SEQUENCE OF 81-400 FROM N.A., AND PARTIAL SEQUENCE OF 63-103.
RX   MEDLINE=90386637; PubMed=2403355;
RA   Sejima H., Hayashi T., Deyashiki Y., Nishioka J., Suzuki K.;
RT   "Primary structure of vitamin K-dependent human protein Z.";
RL   Biochem. Biophys. Res. Commun. 171:661-668(1990).
RN   [4]
RP   STRUCTURE OF CARBOHYDRATE ON SER-93.
RX   MEDLINE=90062160; PubMed=2511201;
RA   Nishimura H., Kawabata S., Kisiel W., Hase S., Ikenaka T., Takao T.,
RA   Shimonishi Y., Iwanaga S.;
RT   "Identification of a disaccharide (Xyl-Glc) and a trisaccharide
RT   (Xyl2-Glc) O-glycosidically linked to a serine residue in the first
RT   epidermal growth factor-like domain of human factors VII and IX and
RT   protein Z and bovine protein Z.";
RL   J. Biol. Chem. 264:20320-20325(1989).
RN   [5]
RP   STRUCTURE OF CARBOHYDRATE ON SER-93.
RX   MEDLINE=91344709; PubMed=2129367;
RA   Iwanaga S., Nishimura H., Kawabata S., Kisiel W., Hase S., Ikenaka T.;
RT   "A new trisaccharide sugar chain linked to a serine residue in the
RT   first EGF-like domain of clotting factors VII and IX and protein Z.";
RL   Adv. Exp. Med. Biol. 281:121-131(1990).
CC   -!- FUNCTION: APPEARS TO ASSIST HEMOSTASIS BY BINDING THROMBIN AND
CC       PROMOTING ITS ASSOCIATION WITH PHOSPHOLIPID VESICLES.
CC   -!- SUBCELLULAR LOCATION: Secreted.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PLASMA.
CC   -!- SIMILARITY: ALTHOUGH HOMOLOGOUS WITH THE VITAMIN K-DEPENDENT
CC       CLOTTING FACTORS, IT HAS LOST TWO OF THE ESSENTIAL CATALYTIC
CC       RESIDUES AND HAS NO ENZYMATIC ACTIVITY.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; M55670; AAA36500.1; -.
DR   EMBL; M55671; AAA36501.1; -.
DR   EMBL; AB033749; BAA85763.1; -.
DR   EMBL; AB033749; BAA85764.1; -.
DR   EMBL; M59303; AAA36499.1; -.
DR   PIR; A36244; A36244.
DR   HSSP; P00740; 1CFH.
DR   MEROPS; S01.979; -.
DR   GlycoSuiteDB; P22891; -.
DR   MIM; 176895; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR002383; GLA_blood.
DR   InterPro; IPR001254; Trypsin.
DR   InterPro; IPR000294; VitK_dep_GLA.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00594; gla; 1.
DR   Pfam; PF00089; trypsin; 1.
DR   PRINTS; PR00001; GLABLOOD.
```

```
DR   SMART; SM00181; EGF; 2.
DR   SMART; SM00069; GLA; 1.
DR   SMART; SM00020; Tryp_SPc; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01186; EGF_2; 2.
DR   PROSITE; PS00011; GLU_CARBOXYLATION; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
KW   Plasma; Glycoprotein; Gamma-carboxyglutamic acid; Hydroxylation;
KW   Calcium; Serine protease homolog; Vitamin K; EGF-like domain; Signal;
KW   Alternative splicing.
FT   SIGNAL       1     23
FT   PROPEP      24     40
FT   CHAIN       41    400       VITAMIN K-DEPENDENT PROTEIN Z.
FT   DOMAIN      87    123       EGF-LIKE 1.
FT   DOMAIN     125    166       EGF-LIKE 2.
FT   DOMAIN     175    400       SERINE PROTEASE.
FT   MOD_RES     47     47       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     48     48       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     51     51       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     55     55       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     57     57       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     60     60       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     61     61       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     66     66       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     67     67       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     70     70       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     73     73       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     75     75       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES     80     80       GAMMA-CARBOXYGLUTAMIC ACID.
FT   MOD_RES    104    104       HYDROXYLATION (BY SIMILARITY).
FT   DISULFID    91    102       BY SIMILARITY.
FT   DISULFID    96    111       BY SIMILARITY.
FT   DISULFID   113    122       BY SIMILARITY.
FT   DISULFID   129    141       BY SIMILARITY.
FT   DISULFID   137    150       BY SIMILARITY.
FT   DISULFID   152    165       BY SIMILARITY.
FT   CARBOHYD    93     93       O-LINKED (GLC. . .).
FT   CARBOHYD    99     99       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   112    112       O-LINKED (POTENTIAL).
FT   CARBOHYD   225    225       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   233    233       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   236    236       O-LINKED (POTENTIAL).
FT   CARBOHYD   306    306       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   315    315       O-LINKED (POTENTIAL).
FT   CARBOHYD   332    332       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    24     24       V -> ATSLKERHGLHSDSACTGVQESL (IN ISOFORM
FT                                2).
SQ   SEQUENCE   400 AA;  44744 MW;  7EBD2DCC48860268 CRC64;


  Query Match           45.7%;  Score 47.5;  DB 1;  Length 400;
  Best Local Similarity  50.0%;  Pred. No. 2.3;
  Matches    9;  Conservative    3;  Mismatches    3;  Indels    3;  Gaps    1;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||:| |: |:   | |||
Db     103 QDSIWGYTCT---CSPGY 117


RESULT  11
CCAE_RAT
ID   CCAE_RAT       STANDARD;      PRT;  2222 AA.
AC   Q07652;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Voltage-dependent R-type calcium channel alpha-1E subunit (Calcium
DE   channel, L type, alpha-1 polypeptide, isoform 6) (RBE-II) (RBE2)
DE   (Brain calcium channel II) (BII).
GN   CACNA1E OR CACNL1A6 OR CACH6.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SPRAGUE-DAWLEY; TISSUE=Brain;
RX   MEDLINE=93262464; PubMed=8388125;
RA   Soong T.W., Stea A., Hodson C.D., Dubel S.J., Vincent S.R.,
RA   Snutch T.P.;
RT   "Structure and functional expression of a member of the low voltage-
RT   activated calcium channel family.";
RL   Science 260:1133-1136(1993).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1E
CC       GIVES RISE TO R-TYPE CALCIUM CURRENTS. R-TYPE CALCIUM CHANNELS
CC       BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED
CC       BY NICKEL, AND PARTIALLY BY OMEGA-AGATOXIN-IIIA (OMEGA-AGA-IIIA).
CC       THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), OMEGA-
CC       CONOTOXIN-GVIA (OMEGA-CTX-GVIA), AND OMEGA-AGATOXIN-IVA (OMEGA-
CC       AGA-IVA). CALCIUM CHANNELS CONTANING ALPHA-1E SUBUNIT COULD BE
CC       INVOLVED IN THE MODULATION OF FIRING PATTERNS OF NEURONS WHICH IS
CC       IMPORTANT FOR INFORMATION PROCESSING.
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN CENTRAL NERVOUS SYSTEM AND IN
CC       INSULINOMA.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L15453; AAA40855.1; -.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation.
FT   REPEAT       27    305       I.
FT   REPEAT      413    657       II.
FT   REPEAT     1092   1378       III.
FT   REPEAT     1415   1678       IV.
FT   DOMAIN        1     40       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     41     59       S1 OF REPEAT I (POTENTIAL).
FT   DOMAIN       60     78       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM     79     97       S2 OF REPEAT I (POTENTIAL).
FT   DOMAIN       98    109       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    110    124       S3 OF REPEAT I (POTENTIAL).
FT   DOMAIN      125    136       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    137    156       S4 OF REPEAT I (POTENTIAL).
FT   DOMAIN      157    174       CYTOPLASMIC (POTENTIAL).
```

```
FT   TRANSMEM    175    195       S5 OF REPEAT I (POTENTIAL).
FT   DOMAIN      196    277       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    278    301       S6 OF REPEAT I (POTENTIAL).
FT   DOMAIN      302    427       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    428    447       S1 OF REPEAT II (POTENTIAL).
FT   DOMAIN      448    460       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    461    480       S2 OF REPEAT II (POTENTIAL).
FT   DOMAIN      481    489       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    490    508       S3 OF REPEAT II (POTENTIAL).
FT   DOMAIN      509    518       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    519    537       S4 OF REPEAT II (POTENTIAL).
FT   DOMAIN      538    556       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    557    576       S5 OF REPEAT II (POTENTIAL).
FT   DOMAIN      577    629       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    630    654       S6 OF REPEAT II (POTENTIAL).
FT   DOMAIN      655   1100       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1101   1117       S1 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1118   1141       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1142   1161       S2 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1162   1169       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1170   1192       S3 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1193   1206       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1207   1224       S4 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1225   1243       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1244   1263       S5 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1264   1350       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1351   1374       S6 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1375   1431       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1432   1450       S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1451   1467       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1468   1485       S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1486   1493       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1494   1512       S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1513   1523       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1524   1542       S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1543   1561       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1562   1581       S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1582   1650       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1651   1676       S6 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1677   2222       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      667    672       POLY-GLU.
FT   DOMAIN      699    704       POLY-ARG.
FT   DOMAIN      718    723       POLY-ARG.
FT   DOMAIN     1058   1064       POLY-GLU.
FT   DOMAIN     1180   1183       POLY-VAL.
FT   DOMAIN     2193   2196       POLY-ARG.
FT   DOMAIN      325    342       BINDING TO THE BETA SUBUNIT (BY
FT                                SIMILARITY).
FT   SITE        260    260       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE        608    608       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1324   1324       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1615   1615       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   CA_BIND     377    388       BY SIMILARITY.
FT   MOD_RES    1686   1686       PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND    1704   1715       BY SIMILARITY.
FT   CARBOHYD    205    205       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1518   1518       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1523   1523       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1641   1641       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2222 AA;  252114 MW;  DF6452A2175CEB19 CRC64;

  Query Match             45.2%;  Score 47;  DB 1;  Length 2222;
  Best Local Similarity   58.3%;  Pred. No. 13;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 HPCSXXGCRPGY 18
           |||   ||  ||
Db     217 HPCGWQGCPAGY 228
```

```
RESULT  12
CCAE_RABIT
ID   CCAE_RABIT     STANDARD;      PRT;  2259 AA.
AC   Q02343; Q02344;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Voltage-dependent R-type calcium channel alpha-1E subunit (Calcium
DE   channel, L type, alpha-1 polypeptide, isoform 6) (Brain calcium
DE   channel II) (BII).
GN   CACNA1E OR CACNL1A6 OR CACH6.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92354772; PubMed=1379552;
RA   Niidome T., Kim M.S., Friedrich T., Mori Y.;
RT   "Molecular cloning and characterization of a novel calcium channel
RT   from rabbit brain.";
RL   FEBS Lett. 308:7-13(1992).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1E
CC       GIVES RISE TO R-TYPE CALCIUM CURRENTS. R-TYPE CALCIUM CHANNELS
CC       BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED
CC       BY NICKEL, AND PARTIALLY BY OMEGA-AGATOXIN-IIIA (OMEGA-AGA-IIIA).
CC       THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), OMEGA-
CC       CONOTOXIN-GVIA (OMEGA-CTX-GVIA), AND OMEGA-AGATOXIN-IVA (OMEGA-
CC       AGA-IVA). CALCIUM CHANNELS CONTANING ALPHA-1E SUBUNIT COULD BE
CC       INVOLVED IN THE MODULATION OF FIRING PATTERNS OF NEURONS WHICH IS
CC       IMPORTANT FOR INFORMATION PROCESSING.
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; BII-1 (SHOWN HERE) AND BII-2;
CC       ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: ABUNDANT IN THE CEREBRAL CORTEX, HIPPOCAMPUS,
CC       AND CORPUS STRIATUM.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X67855; CAA48040.1; -.
DR   EMBL; X67856; CAA48041.1; -.
DR   PIR; S29236; S29236.
DR   PIR; S29237; S29237.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
```

```
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation; Alternative splicing.
FT   REPEAT       76    354       I.
FT   REPEAT      464    706       II.
FT   REPEAT     1130   1414       III.
FT   REPEAT     1453   1716       IV.
FT   DOMAIN        1     89       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     90    108       S1 OF REPEAT I.
FT   DOMAIN      109    126       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    127    146       S2 OF REPEAT I.
FT   DOMAIN      147    158       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    159    176       S3 OF REPEAT I.
FT   DOMAIN      177    185       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    186    204       S4 OF REPEAT I.
FT   DOMAIN      205    223       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    224    243       S5 OF REPEAT I.
FT   DOMAIN      244    326       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    327    351       S6 OF REPEAT I.
FT   DOMAIN      352    476       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    477    495       S1 OF REPEAT II.
FT   DOMAIN      496    510       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    511    530       S2 OF REPEAT II.
FT   DOMAIN      531    538       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    539    557       S3 OF REPEAT II.
FT   DOMAIN      558    567       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    568    586       S4 OF REPEAT II.
FT   DOMAIN      587    605       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    606    625       S5 OF REPEAT II.
FT   DOMAIN      626    678       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    679    703       S6 OF REPEAT II.
FT   DOMAIN      704   1143       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1144   1162       S1 OF REPEAT III.
FT   DOMAIN     1163   1178       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1179   1198       S2 OF REPEAT III.
FT   DOMAIN     1199   1210       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1211   1229       S3 OF REPEAT III.
FT   DOMAIN     1230   1243       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1244   1262       S4 OF REPEAT III.
FT   DOMAIN     1263   1281       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1282   1301       S5 OF REPEAT III.
FT   DOMAIN     1302   1388       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1389   1413       S6 OF REPEAT III.
FT   DOMAIN     1414   1468       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1469   1487       S1 OF REPEAT IV.
FT   DOMAIN     1488   1502       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1503   1522       S2 OF REPEAT IV.
FT   DOMAIN     1523   1530       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1531   1549       S3 OF REPEAT IV.
FT   DOMAIN     1550   1561       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1562   1580       S4 OF REPEAT IV.
FT   DOMAIN     1581   1599       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1600   1619       S5 OF REPEAT IV.
FT   DOMAIN     1620   1688       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1689   1712       S6 OF REPEAT IV.
FT   DOMAIN     1713   2259       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      716    721       POLY-GLU.
FT   DOMAIN      748    753       POLY-ARG.
FT   DOMAIN      767    772       POLY-ARG.
FT   DOMAIN     1218   1221       POLY-VAL.
FT   DOMAIN     1976   1979       POLY-SER.
FT   DOMAIN     2231   2235       POLY-ARG.
FT   DOMAIN      374    391       BINDING TO THE BETA SUBUNIT (BY
FT                                SIMILARITY).
FT   SITE        309    309       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE        657    657       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1362   1362       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1653   1653       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   CA_BIND     426    437       BY SIMILARITY.
FT   MOD_RES    1724   1724       PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND    1742   1753       BY SIMILARITY.
FT   CARBOHYD    254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1556   1556       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1561   1561       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC   2101   2259       HSRRQLPPVPPKPRPLLSYSSLKQQPSNFSPPADGSQGGSL
FT                                LASPALESAQVGLPESSDSPRRAQGSHASPQRYISEPYLAL
FT                                HEDSHASDCGEEETLTFEAAVATSLGRSNTIGSAPPLRHSW
FT                                QMPNGHYRRRRRGGPGAGALCGAVGDLLSDTEEDKC -> Q
FT                                QWGPQEEGVLLHPHQGCGWPCRDRRWMPGRRGWSGEKSHSP
FT                                LPHCGRDSTGGAGQGPPRYCGSGAGDAGGTCDSLSP (IN
FT                                ISOFORM BII-2).
SQ   SEQUENCE   2259 AA;  254250 MW;  E4A757076E38779E CRC64;


  Query Match             45.2%;  Score 47;  DB 1;  Length 2259;
  Best Local Similarity   58.3%;  Pred. No. 14;
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 HPCSXXGCRPGY 18
           |||   ||  ||
Db     266 HPCGVQGCPAGY 277


RESULT 13
CCAE_MOUSE
ID   CCAE_MOUSE     STANDARD;      PRT;  2272 AA.
AC   Q61290;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   30-MAY-2000 (Rel. 39, Last annotation update)
DE   Voltage-dependent R-type calcium channel alpha-1E subunit (Calcium
DE   channel, L type, alpha-1 polypeptide, isoform 6) (Brain calcium
DE   channel II) (BII).
GN   CACNA1E OR CCHRA1 OR CACNL1A6 OR CACH6.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=Brain;
RX   MEDLINE=94350992; PubMed=8071363;
RA   Williams M.E., Marubio L.M., Deal C.R., Hans M., Brust P.F.,
RA   Philipson L.H., Miller R.J., Johnson E.C., Harpold M.M., Ellis S.B.;
RT   "Structure and functional characterization of neuronal alpha 1E
RT   calcium channel subtypes.";
RL   J. Biol. Chem. 269:22347-22357(1994).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1E
CC       GIVES RISE TO R-TYPE CALCIUM CURRENTS. R-TYPE CALCIUM CHANNELS
CC       BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED
CC       BY NICKEL, AND PARTIALLY BY OMEGA-AGATOXIN-IIIA (OMEGA-AGA-IIIA).
CC       THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), OMEGA-
CC       CONOTOXIN-GVIA (OMEGA-CTX-GVIA), AND OMEGA-AGATOXIN-IVA (OMEGA-
CC       AGA-IVA). CALCIUM CHANNELS CONTANING ALPHA-1E SUBUNIT COULD BE
CC       INVOLVED IN THE MODULATION OF FIRING PATTERNS OF NEURONS WHICH IS
CC       IMPORTANT FOR INFORMATION PROCESSING.
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONAL TISSUES, RETINA, SPLEEN,
```

```
CC       AND PANCREATIC ISLET CELLS.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L29346; AAA59206.1; -.
DR   MGD; MGI:106217; Cacnale.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation.
FT   REPEAT       77    355       I.
FT   REPEAT      463    707       II.
FT   REPEAT     1143   1429       III.
FT   REPEAT     1466   1729       IV.
FT   DOMAIN        1     90       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     91    109       S1 OF REPEAT I (POTENTIAL).
FT   DOMAIN      110    128       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    129    147       S2 OF REPEAT I (POTENTIAL).
FT   DOMAIN      148    159       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    160    174       S3 OF REPEAT I (POTENTIAL).
FT   DOMAIN      175    186       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    187    206       S4 OF REPEAT I (POTENTIAL).
FT   DOMAIN      207    224       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    225    245       S5 OF REPEAT I (POTENTIAL).
FT   DOMAIN      246    327       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    328    351       S6 OF REPEAT I (POTENTIAL).
FT   DOMAIN      352    477       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    478    497       S1 OF REPEAT II (POTENTIAL).
FT   DOMAIN      498    510       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    511    530       S2 OF REPEAT II (POTENTIAL).
FT   DOMAIN      531    539       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    540    558       S3 OF REPEAT II (POTENTIAL).
FT   DOMAIN      559    568       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    569    587       S4 OF REPEAT II (POTENTIAL).
FT   DOMAIN      588    606       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    607    626       S5 OF REPEAT II (POTENTIAL).
FT   DOMAIN      627    679       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    680    704       S6 OF REPEAT II (POTENTIAL).
FT   DOMAIN      705   1150       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1151   1167       S1 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1168   1191       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1192   1211       S2 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1212   1219       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1220   1242       S3 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1243   1256       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1257   1274       S4 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1275   1293       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1294   1313       S5 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1314   1400       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1401   1424       S6 OF REPEAT III (POTENTIAL).
FT   DOMAIN     1425   1481       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1482   1500       S1 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1501   1515       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1516   1535       S2 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1536   1543       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1544   1562       S3 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1563   1573       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1574   1592       S4 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1593   1611       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1612   1631       S5 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1632   1700       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1701   1726       S6 OF REPEAT IV (POTENTIAL).
FT   DOMAIN     1727   2272       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      717    722       POLY-GLU.
FT   DOMAIN      751    754       POLY-ARG.
FT   DOMAIN      770    773       POLY-ARG.
FT   DOMAIN     1108   1112       POLY-GLU.
FT   DOMAIN     1115   1118       POLY-LYS.
FT   DOMAIN     1231   1234       POLY-VAL.
FT   DOMAIN     2244   2247       POLY-ARG.
FT   DOMAIN      375    392       BINDING TO THE BETA SUBUNIT (BY
FT                                SIMILARITY).
FT   SITE        310    310       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE        658    658       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1375   1375       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1666   1666       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   CA_BIND     427    438       BY SIMILARITY.
FT   MOD_RES    1737   1737       PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND    1755   1766       BY SIMILARITY.
FT   CARBOHYD    255    255       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1569   1569       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1692   1692       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   2272 AA;  257233 MW;  70D9200B9E0C87A1 CRC64;
```

Query Match 45.2%; Score 47; DB 1; Length 2272;
Best Local Similarity 58.3%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy       7 HPCSXXGCRPGY 18
           |||   ||  ||
Db     267 HPCGVQGCPAGY 278
```

```
RESULT  14
CCAE_HUMAN
ID   CCAE_HUMAN     STANDARD;      PRT;  2312 AA.
AC   Q15878; Q14581; Q14580;
DT   15-JUL-1999 (Rel. 38, Created)
DT   15-JUL-1999 (Rel. 38, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Voltage-dependent R-type calcium channel alpha-1E subunit (Calcium
DE   channel, L type, alpha-1 polypeptide, isoform 6) (Brain calcium
DE   channel II) (BII).
GN   CACNA1E OR CACNL1A6 OR CACH6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95236033; PubMed=7536609;
RA   Schneider T., Wei X., Olcese R., Costantin J.L., Neely A., Palade P.,
RA   Perez-Reyes E., Qin N., Zhou J., Crawford G.D., Smith R.G.,
RA   Appel S.H., Stefani E., Birnbaumer M.;
RT   "Molecular analysis and functional expression of the human type E
RT   neuronal Ca2+ channel alpha 1 subunit.";
RL   Recept. Channels 2:255-270(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=94350992; PubMed=8071363;
RA   Williams M.E., Marubio L.M., Deal C.R., Hans M., Brust P.F.,
```

```
RA   Philipson L.H., Miller R.J., Johnson E.C., Harpold M.M., Ellis S.B.;
RT   "Structure and functional characterization of neuronal alpha 1E
RT   calcium channel subtypes.";
RL   J. Biol. Chem. 269:22347-22357(1994).
CC   -!- FUNCTION: VOLTAGE-SENSITIVE CALCIUM CHANNELS (VSCC) MEDIATE THE
CC       ENTRY OF CALCIUM IONS INTO EXCITABLE CELLS AND ARE ALSO INVOLVED
CC       IN A VARIETY OF CALCIUM-DEPENDENT PROCESSES, INCLUDING MUSCLE
CC       CONTRACTION, HORMONE OR NEUROTRANSMITTER RELEASE, GENE EXPRESSION,
CC       CELL MOTILITY, CELL DIVISION AND CELL DEATH. THE ISOFORM ALPHA-1E
CC       GIVES RISE TO R-TYPE CALCIUM CURRENTS. R-TYPE CALCIUM CHANNELS
CC       BELONG TO THE "HIGH-VOLTAGE ACTIVATED" (HVA) GROUP AND ARE BLOCKED
CC       BY NICKEL, AND PARTIALLY BY OMEGA-AGATOXIN-IIIA (OMEGA-AGA-IIIA).
CC       THEY ARE HOWEVER INSENSITIVE TO DIHYDROPYRIDINES (DHP), OMEGA-
CC       CONOTOXIN-GVIA (OMEGA-CTX-GVIA), AND OMEGA-AGATOXIN-IVA (OMEGA-
CC       AGA-IVA). CALCIUM CHANNELS CONTANING ALPHA-1E SUBUNIT COULD BE
CC       INVOLVED IN THE MODULATION OF FIRING PATTERNS OF NEURONS WHICH IS
CC       IMPORTANT FOR INFORMATION PROCESSING.
CC   -!- SUBUNIT: VOLTAGE-DEPENDENT CALCIUM CHANNELS ARE MULTISUBUNIT
CC       COMPLEXES, CONSISTING OF ALPHA-1, ALPHA-2, BETA AND DELTA SUBUNITS
CC       IN A 1:1:1:1 RATIO. THE CHANNEL ACTIVITY IS DIRECTED BY THE PORE-
CC       FORMING AND VOLTAGE-SENSITIVE ALPHA-1 SUBUNIT. IN MANY CASES, THIS
CC       SUBUNIT IS SUFFICIENT TO GENERATE VOLTAGE-SENSITIVE CALCIUM
CC       CHANNEL ACTIVITY. THE AUXILIARY SUBUNITS BETA AND ALPHA-2/DELTA
CC       LINKED BY A DISULFIDE BRIDGE REGULATE THE CHANNEL ACTIVITY.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA-1E-1 AND ALPHA-1E-3
CC       (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN NEURONAL TISSUES AND IN KIDNEY.
CC   -!- DOMAIN: EACH OF THE FOUR INTERNAL REPEATS CONTAINS FIVE
CC       HYDROPHOBIC TRANSMEMBRANE SEGMENTS (S1, S2, S3, S5, S6) AND ONE
CC       POSITIVELY CHARGED TRANSMEMBRANE SEGMENT (S4). S4 SEGMENTS
CC       PROBABLY REPRESENT THE VOLTAGE-SENSOR AND ARE CHARACTERIZED BY A
CC       SERIES OF POSITIVELY CHARGED AMINO ACIDS AT EVERY THIRD POSITION.
CC   -!- SIMILARITY: BELONGS TO THE CALCIUM CHANNEL ALPHA-1 SUBUNITS
CC       FAMILY.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; L27745; AAA72125.1; -.
DR   EMBL; L29384; AAA59204.1; -.
DR   EMBL; L29385; AAA59205.1; -.
DR   MIM; 601013; -.
DR   InterPro; IPR002077; Ca_channel.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR000636; Cation_chan_non_lig.
DR   InterPro; IPR001682; Channel_pore_Ca_Na.
DR   Pfam; PF00520; ion_trans; 4.
DR   PRINTS; PR00167; CACHANNEL.
KW   Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW   Calcium channel; Glycoprotein; Repeat; Multigene family;
KW   Calcium-binding; Phosphorylation; Alternative splicing.
FT   REPEAT       76    354       I (BY SIMILARITY).
FT   REPEAT      462    706       II (BY SIMILARITY).
FT   REPEAT     1139   1425       III (BY SIMILARITY).
FT   REPEAT     1462   1725       IV (BY SIMILARITY).
FT   DOMAIN        1     89       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     90    108       S1 OF REPEAT I.
FT   DOMAIN      109    127       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    128    146       S2 OF REPEAT I.
FT   DOMAIN      147    158       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    159    173       S3 OF REPEAT I.
FT   DOMAIN      174    185       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    186    205       S4 OF REPEAT I.
FT   DOMAIN      206    223       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    224    244       S5 OF REPEAT I.
FT   DOMAIN      245    326       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    327    350       S6 OF REPEAT I.
FT   DOMAIN      351    476       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    477    496       S1 OF REPEAT II.
FT   DOMAIN      497    509       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    510    529       S2 OF REPEAT II.
FT   DOMAIN      530    538       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    539    557       S3 OF REPEAT II.
FT   DOMAIN      558    567       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    568    586       S4 OF REPEAT II.
FT   DOMAIN      587    605       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM    606    625       S5 OF REPEAT II.
FT   DOMAIN      626    678       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM    679    703       S6 OF REPEAT II.
FT   DOMAIN      704   1147       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1148   1164       S1 OF REPEAT III.
FT   DOMAIN     1165   1188       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1189   1208       S2 OF REPEAT III.
FT   DOMAIN     1209   1216       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1217   1239       S3 OF REPEAT III.
FT   DOMAIN     1240   1253       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1254   1271       S4 OF REPEAT III.
FT   DOMAIN     1272   1290       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1291   1310       S5 OF REPEAT III.
FT   DOMAIN     1311   1397       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1398   1421       S6 OF REPEAT III.
FT   DOMAIN     1422   1478       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1479   1497       S1 OF REPEAT IV.
FT   DOMAIN     1498   1512       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1513   1532       S2 OF REPEAT IV.
FT   DOMAIN     1533   1540       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1541   1559       S3 OF REPEAT IV.
FT   DOMAIN     1560   1570       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1571   1589       S4 OF REPEAT IV.
FT   DOMAIN     1590   1608       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM   1609   1628       S5 OF REPEAT IV.
FT   DOMAIN     1629   1697       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1698   1723       S6 OF REPEAT IV.
FT   DOMAIN     1724   2312       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN      716    721       POLY-GLU.
FT   DOMAIN      748    753       POLY-ARG.
FT   DOMAIN      767    772       POLY-ARG.
FT   DOMAIN     1227   1230       POLY-VAL.
FT   DOMAIN     2283   2287       POLY-ARG.
FT   DOMAIN      374    391       BINDING TO THE BETA SUBUNIT (BY
FT                                SIMILARITY).
FT   SITE        309    309       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE        657    657       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1371   1371       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   SITE       1662   1662       CALCIUM ION SELECTIVITY AND PERMEABILITY
FT                                (BY SIMILARITY).
FT   CA_BIND     426    437       BY SIMILARITY.
FT   MOD_RES    1733   1733       PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT   CA_BIND    1751   1762       BY SIMILARITY.
FT   CARBOHYD    254    254       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1565   1565       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1570   1570       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    748    766       MISSING (IN ISOFORM ALPHA-1E-1).
FT   CONFLICT    648    648       I -> M (IN REF. 2).
FT   CONFLICT    836    837       WP -> LAL (IN REF. 2).
FT   CONFLICT   1954   1954       T -> A (IN REF. 2).
FT   CONFLICT   1966   2008       MISSING (IN REF. 2).
FT   CONFLICT   2076   2076       R -> P (IN REF. 2).
FT   CONFLICT   2083   2083       G -> R (IN REF. 2).
FT   CONFLICT   2205   2205       C -> W (IN REF. 2).
FT   CONFLICT   2218   2218       S -> R (IN REF. 2).
FT   CONFLICT   2244   2244       G -> V (IN REF. 2).
SQ   SEQUENCE   2312 AA;  261727 MW;  633ED3EFD407D65E CRC64;

 Query Match             45.2%;  Score 47;  DB 1;  Length 2312;
 Best Local Similarity   58.3%;  Pred. No. 14;
```

```
  Matches    7;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       7 HPCSXXGCRPGY 18
           |||   ||  ||
Db     266 HPCGVQGCPAGY 277


RESULT  15
NTC3_MOUSE
ID   NTC3_MOUSE     STANDARD;      PRT;  2318 AA.
AC   Q61982;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurogenic locus notch 3 protein.
GN   NOTCH3.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ICR.X SWISS WEBSTER;
RX   MEDLINE=95001556; PubMed=7918097;
RA   Lardelli M., Dalstrand J., Lendahl U.;
RT   "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT   growth factor-repeats and is expressed in proliferating
RT   neuroepithelium.";
RL   Mech. Dev. 46:123-136(1994).
CC   -!- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC       VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC       DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC   -!- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC   -!- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC   -!- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; X74760; CAA52776.1; -.
DR   HSSP; P00740; 1IXA.
DR   MGD; MGI:99460; Notch3.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR000800; Notch.
DR   Pfam; PF00023; ank; 6.
DR   Pfam; PF00008; EGF; 34.
DR   Pfam; PF00066; notch; 3.
DR   PRINTS; PR00010; EGFBLOOD.
DR   PRINTS; PR01452; NOTCH.
DR   SMART; SM00248; ANK; 5.
DR   SMART; SM00179; EGF_CA; 19.
DR   SMART; SM00001; EGF_like; 13.
DR   SMART; SM00004; NL; 3.
DR   PROSITE; PS50088; ANK_REPEAT; 4.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS00010; ASX_HYDROXYL; 18.
DR   PROSITE; PS00022; EGF_1; 33.
DR   PROSITE; PS01186; EGF_2; 27.
DR   PROSITE; PS01187; EGF_CA; 17.
KW   Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW   ANK repeat; Glycoprotein.

FT   DOMAIN       1   1643   EXTRACELLULAR.
FT   TRANSMEM  1644   1664   POTENTIAL.
FT   DOMAIN    1665   2318   CYTOPLASMIC.
FT   DOMAIN    2242   2261   PEST.
FT   DOMAIN      39     78   EGF-LIKE 1.
FT   DOMAIN      79    119   EGF-LIKE 2.
FT   DOMAIN     120    157   EGF-LIKE 3.
FT   DOMAIN     159    196   EGF-LIKE 4, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     198    235   EGF-LIKE 5.
FT   DOMAIN     237    273   EGF-LIKE 6, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     275    313   EGF-LIKE 7.
FT   DOMAIN     315    351   EGF-LIKE 8, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     352    390   EGF-LIKE 9.
FT   DOMAIN     392    430   EGF-LIKE 10, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     432    468   EGF-LIKE 11, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     470    506   EGF-LIKE 12, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     508    544   EGF-LIKE 13, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     546    581   EGF-LIKE 14, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     583    619   EGF-LIKE 15, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     621    656   EGF-LIKE 16, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     658    694   EGF-LIKE 17, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     696    731   EGF-LIKE 18.
FT   DOMAIN     735    771   EGF-LIKE 19.
FT   DOMAIN     772    809   EGF-LIKE 20.
FT   DOMAIN     811    848   EGF-LIKE 21, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     850    886   EGF-LIKE 22, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     888    923   EGF-LIKE 23, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     925    961   EGF-LIKE 24.
FT   DOMAIN     963    999   EGF-LIKE 25.
FT   DOMAIN    1001   1035   EGF-LIKE 26.
FT   DOMAIN    1037   1083   EGF-LIKE 27.
FT   DOMAIN    1085   1121   EGF-LIKE 28.
FT   DOMAIN    1123   1159   EGF-LIKE 29, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1161   1204   EGF-LIKE 30, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN    1206   1245   EGF-LIKE 31.
FT   DOMAIN    1247   1288   EGF-LIKE 32.
FT   DOMAIN    1290   1326   EGF-LIKE 33.
FT   DOMAIN    1336   1374   EGF-LIKE 34.
FT   REPEAT    1388   1428   LIN/NOTCH 1.
FT   REPEAT    1429   1467   LIN/NOTCH 2.
FT   REPEAT    1468   1503   LIN/NOTCH 3.
FT   REPEAT    1839   1868   ANK 1.
FT   REPEAT    1872   1902   ANK 2.
FT   REPEAT    1906   1935   ANK 3.
FT   REPEAT    1939   1968   ANK 4.
FT   REPEAT    1972   2001   ANK 5.
FT   DISULFID    43     55   BY SIMILARITY.
FT   DISULFID    49     66   BY SIMILARITY.
FT   DISULFID    68     77   BY SIMILARITY.
FT   DISULFID    83     94   BY SIMILARITY.
FT   DISULFID    88    107   BY SIMILARITY.
FT   DISULFID   109    118   BY SIMILARITY.
FT   DISULFID   124    135   BY SIMILARITY.
FT   DISULFID   129    145   BY SIMILARITY.
FT   DISULFID   147    156   BY SIMILARITY.
FT   DISULFID   163    175   BY SIMILARITY.
FT   DISULFID   169    184   BY SIMILARITY.
FT   DISULFID   186    195   BY SIMILARITY.
FT   DISULFID   202    213   BY SIMILARITY.
FT   DISULFID   207    223   BY SIMILARITY.
FT   DISULFID   225    234   BY SIMILARITY.
FT   DISULFID   241    252   BY SIMILARITY.
FT   DISULFID   246    261   BY SIMILARITY.
FT   DISULFID   263    272   BY SIMILARITY.
FT   DISULFID   279    292   BY SIMILARITY.
FT   DISULFID   286    301   BY SIMILARITY.
FT   DISULFID   303    312   BY SIMILARITY.
FT   DISULFID   319    330   BY SIMILARITY.
FT   DISULFID   324    339   BY SIMILARITY.
FT   DISULFID   341    350   BY SIMILARITY.
FT   DISULFID   356    367   BY SIMILARITY.
FT   DISULFID   361    378   BY SIMILARITY.
FT   DISULFID   380    389   BY SIMILARITY.
```

FT	DISULFID	396	409	BY SIMILARITY.
FT	DISULFID	403	418	BY SIMILARITY.
FT	DISULFID	420	429	BY SIMILARITY.
FT	DISULFID	436	447	BY SIMILARITY.
FT	DISULFID	441	456	BY SIMILARITY.
FT	DISULFID	458	467	BY SIMILARITY.
FT	DISULFID	474	485	BY SIMILARITY.
FT	DISULFID	479	494	BY SIMILARITY.
FT	DISULFID	496	505	BY SIMILARITY.
FT	DISULFID	512	523	BY SIMILARITY.
FT	DISULFID	517	532	BY SIMILARITY.
FT	DISULFID	534	543	BY SIMILARITY.
FT	DISULFID	550	560	BY SIMILARITY.
FT	DISULFID	555	569	BY SIMILARITY.
FT	DISULFID	571	580	BY SIMILARITY.
FT	DISULFID	587	598	BY SIMILARITY.
FT	DISULFID	592	607	BY SIMILARITY.
FT	DISULFID	609	618	BY SIMILARITY.
FT	DISULFID	625	635	BY SIMILARITY.
FT	DISULFID	630	644	BY SIMILARITY.
FT	DISULFID	646	655	BY SIMILARITY.
FT	DISULFID	662	673	BY SIMILARITY.
FT	DISULFID	667	682	BY SIMILARITY.
FT	DISULFID	684	693	BY SIMILARITY.
FT	DISULFID	700	710	BY SIMILARITY.
FT	DISULFID	705	719	BY SIMILARITY.
FT	DISULFID	721	730	BY SIMILARITY.
FT	DISULFID	739	750	BY SIMILARITY.
FT	DISULFID	744	759	BY SIMILARITY.
FT	DISULFID	761	770	BY SIMILARITY.
FT	DISULFID	776	787	BY SIMILARITY.
FT	DISULFID	781	797	BY SIMILARITY.
FT	DISULFID	799	808	BY SIMILARITY.
FT	DISULFID	815	827	BY SIMILARITY.
FT	DISULFID	821	836	BY SIMILARITY.
FT	DISULFID	838	847	BY SIMILARITY.
FT	DISULFID	854	865	BY SIMILARITY.
FT	DISULFID	859	874	BY SIMILARITY.
FT	DISULFID	876	885	BY SIMILARITY.
FT	DISULFID	892	902	BY SIMILARITY.
FT	DISULFID	897	911	BY SIMILARITY.
FT	DISULFID	913	922	BY SIMILARITY.
FT	DISULFID	929	940	BY SIMILARITY.
FT	DISULFID	934	949	BY SIMILARITY.
FT	DISULFID	951	960	BY SIMILARITY.
FT	DISULFID	967	978	BY SIMILARITY.
FT	DISULFID	972	987	BY SIMILARITY.
FT	DISULFID	989	998	BY SIMILARITY.
FT	DISULFID	1005	1016	BY SIMILARITY.
FT	DISULFID	1010	1023	BY SIMILARITY.
FT	DISULFID	1025	1034	BY SIMILARITY.
FT	DISULFID	1041	1062	BY SIMILARITY.
FT	DISULFID	1056	1071	BY SIMILARITY.
FT	DISULFID	1073	1082	BY SIMILARITY.
FT	DISULFID	1089	1100	BY SIMILARITY.
FT	DISULFID	1094	1109	BY SIMILARITY.
FT	DISULFID	1111	1120	BY SIMILARITY.
FT	DISULFID	1127	1138	BY SIMILARITY.
FT	DISULFID	1132	1147	BY SIMILARITY.
FT	DISULFID	1149	1158	BY SIMILARITY.
FT	DISULFID	1165	1183	BY SIMILARITY.
FT	DISULFID	1177	1192	BY SIMILARITY.
FT	DISULFID	1194	1203	BY SIMILARITY.

Query Match 44.7%; Score 46.5; DB 1; Length 2318;
Best Local Similarity 40.9%; Pred. No. 17;
Matches 9; Conservative 1; Mismatches 5; Indels 7; Gaps 1;

```
Qy       4 IHGHPCS-------XXGCRPGY 18
           :|| |||          | |||
Db     130 VHGAPCSVGPDGRFACACPPGY 151
```

Search completed: August 26, 2002, 13:35:49
Job time: 363 sec

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:35:10 ; Search time 35.4 Seconds
(without alignments)
87.964 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	61	58.7	990	4	Q15206	Q15206 homo sapien
2	61	58.7	1218	4	Q05331	Q05331 homo sapien
3	54	51.9	465	4	Q03838	Q03838 homo sapien
4	54	51.9	591	4	Q01720	Q01720 homo sapien
5	54	51.9	687	4	Q9H4U2	Q9h4u2 homo sapien
6	53	51.0	797	4	Q16824	Q16824 homo sapien
7	53	51.0	798	4	Q9H4U3	Q9h4u3 homo sapien
8	53	51.0	1084	4	Q01212	Q01212 homo sapien
9	51	49.0	377	12	Q89164	Q89164 variola vir
10	51	49.0	377	12	Q85389	Q85389 variola maj
11	51	49.0	377	12	O93122	O93122 vaccinia vi
12	50.5	48.6	209	11	Q9D8P5	Q9d8p5 mus musculu
13	50.5	48.6	209	11	Q91VN7	Q91vn7 mus musculu
14	50	48.1	397	1	Q9HH11	Q9hh11 thermococcu
15	49	47.1	322	4	O75370	O75370 homo sapien
16	47.5	45.7	921	5	Q969A3	Q969a3 branchiosto
17	47	45.2	372	3	Q9Y850	Q9y850 kluyveromyc
18	47	45.2	2162	4	Q9NYZ5	Q9nyz5 homo sapien
19	47	45.2	2224	4	Q9NYZ6	Q9nyz6 homo sapien
20	47	45.2	2295	11	Q923K6	Q923k6 rattus norv
21	46	44.2	221	5	Q18883	Q18883 caenorhabdi
22	46	44.2	397	17	Q9UYI3	Q9uyi3 pyrococcus
23	46	44.2	2439	2	Q9KIZ5	Q9kiz5 polyangium
24	45.5	43.8	274	5	Q9W2W3	Q9w2w3 drosophila
25	45.5	43.8	815	4	Q96JS2	Q96js2 homo sapien
26	45	43.3	174	10	Q9SC35	Q9sc35 pisum sativ
27	45	43.3	1234	2	Q9AES2	Q9aes2 clostridium
28	44.5	42.8	3767	5	Q9UA13	Q9ua13 caenorhabdi
29	44.5	42.8	4601	5	Q9V383	Q9v383 drosophila
30	44	42.3	397	17	O58261	O58261 pyrococcus
31	44	42.3	529	5	Q9VQ48	Q9vq48 drosophila
32	44	42.3	830	4	Q14162	Q14162 homo sapien
33	44	42.3	830	4	O43701	O43701 homo sapien
34	44	42.3	942	16	Q9Z6V1	Q9z6v1 chlamydia p
35	44	42.3	1224	5	O96209	O96209 plasmodium
36	43.5	41.8	153	4	O75095	O75095 homo sapien
37	43.5	41.8	2192	5	O01768	O01768 caenorhabdi
38	43.5	41.8	2319	11	Q9R172	Q9r172 rattus norv
39	43	41.3	59	5	Q9XV23	Q9xv23 caenorhabdi
40	43	41.3	126	2	Q9L192	Q91192 streptomyce
41	43	41.3	374	5	Q9VFA2	Q9vfa2 drosophila
42	43	41.3	603	2	Q938F6	Q938f6 rhodococcus
43	43	41.3	976	13	O73878	O73878 brachydanio
44	42.5	40.9	1488	5	Q20294	Q20294 caenorhabdi
45	42.5	40.9	2447	13	O13149	O13149 fugu rubrip

ALIGNMENTS

RESULT 1
Q15206
ID Q15206 PRELIMINARY; PRT; 990 AA.
AC Q15206;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PROFILAGGRIN (FRAGMENT).
GN FLG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=91064347; PubMed=2248957;
RA Gan S.Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;
RT "Organization, structure, and polymorphisms of the human profilaggrin
RT gene [published erratum appears in Biochemistry 1991 Jun
RT 11;30(23):5814].";
RL Biochemistry 29:9432-9440(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=91255199; PubMed=2043621;
RA Gan S.Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;
RT "Organization, structure, and polymorphisms of the human profilaggrin
RT gene.";
RL Biochemistry 30:5814-5814(1991).
DR EMBL; M60494; AAA63244.1; -.
DR InterPro; IPR003303; Filaggrin.
DR PRINTS; PR00487; FILAGGRIN.
FT NON_TER 990 990
SQ SEQUENCE 990 AA; 106453 MW; A8396F10F6A91991 CRC64;

Query Match 58.7%; Score 61; DB 4; Length 990;
Best Local Similarity 66.7%; Pred. No. 0.049;

```
  Matches   12;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPGY 18
          ||||| || |  | | ||
Db    291 QDTIHAHPGSRRGGRHGY 308


RESULT   2
Q05331
ID   Q05331      PRELIMINARY;      PRT;  1218 AA.
AC   Q05331;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FILAGGRIN (PROFILAGGRIN) (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=FORESKIN;
RX   MEDLINE=93109348; PubMed=8417356;
RA   Markova N.G., Marekov L.N., Chipev C.C., Gan S.-Q., Idler W.W.,
RA   Steinert P.M.;
RT   "Profilaggrin is a major epidermal calcium-binding protein.";
RL   Mol. Cell. Biol. 13:613-625(1993).
CC   -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC       DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING
CC       TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC   -!- PTM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC       HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF
CC       324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES.
CC       DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED &
CC       PROTEOLYTICALLY CLEAVED.
CC   -!- POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND
CC       WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FILAGGRIN
CC       REPEATS.
DR   EMBL; M96943; AAA36487.1; -.
DR   HSSP; P02593; 1CDM.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003303; Filaggrin.
DR   InterPro; IPR001751; S100_CaBP.
DR   Pfam; PF00036; efhand; 1.
DR   Pfam; PF01023; S_100; 1.
DR   PRINTS; PR00487; FILAGGRIN.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Phosphorylation; Polyprotein; Developmental protein; Calcium-binding;
KW   Polymorphism.
FT   CA_BIND      19     32       SITE I (BY SIMILARITY).
FT   CA_BIND      62     73       SITE II (BY SIMILARITY).
FT   NON_TER    1218   1218
SQ   SEQUENCE   1218 AA;  133604 MW;  EC195AD5285B19C2 CRC64;


  Query Match             58.7%;  Score 61;  DB 4;  Length 1218;
  Best Local Similarity   66.7%;  Pred. No. 0.06;
  Matches   12;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPGY 18
          ||||| || |  | | ||
Db    513 QDTIHAHPGSRRGGRHGY 530


RESULT   3
Q03838
ID   Q03838      PRELIMINARY;      PRT;  465 AA.
AC   Q03838;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FILAGGRIN (PROFILAGGRIN) (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=91064347; PubMed=2248957;
RA   Gan S.-Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;
RT   "Organization, structure, and polymorphisms of the human profilaggrin
RT   gene.";
RL   Biochemistry 29:9432-9440(1990).
RN   [2]
RP   REVISIONS.
RX   MEDLINE=91255199; PubMed=2043621;
RA   Gan S.-Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;
RT   "Organization, structure, and polymorphisms of the human profilaggrin
RT   gene.";
RL   Biochemistry 30:5814-5814(1991).
CC   -!- FUNCTION: FILAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND
CC       PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE
CC       FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC   -!- POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND
CC       WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FILAGGRIN
CC       REPEATS.
CC   -!- MISCELLANEOUS: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE,
CC       INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM
CC       COPIES OF 324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHYALIN
CC       GRANULES. DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED &
CC       PROTEOLYTICALLY CLEAVED.
DR   EMBL; M60499; AAA63246.1; -.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
FT   NON_TER       1      1
FT   NON_TER     465    465
SQ   SEQUENCE   465 AA;  50280 MW;  C883744C5E134097 CRC64;


  Query Match             51.9%;  Score 54;  DB 4;  Length 465;
  Best Local Similarity   64.7%;  Pred. No. 0.36;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||| ||| |  | | |
Db    291 QDTIRGHPGSSRGGRQG 307


RESULT   4
Q01720
ID   Q01720      PRELIMINARY;      PRT;  591 AA.
AC   Q01720;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   FILAGGRIN PRECURSOR (PROFILAGGRIN) (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=93054736; PubMed=1429717;
RA   Presland R.B., Haydock P.V., Fleckman P., Nirunsuksiri W., Dale B.A.;
RT   "Characterization of the human epidermal profilaggrin gene. Genomic
RT   organization and identification of an S-100-like calcium binding
RT   domain at the amino terminus.";
RL   J. Biol. Chem. 267:23772-23781(1992).
CC   -!- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC       DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE FILAMENTS DURING
CC       TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
```

```
CC   -!- PTM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC       HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES OF
CC       324 AA. THE PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES.
CC       DURING TERMINAL DIFFERENTIATION IT IS DEPHOSPHORYLATED &
CC       PROTEOLYTICALLY CLEAVED.
CC   -!- POLYMORPHISM: A NUMBER OF PROFILAGGRIN ISOFORMS HAVE BEEN FOUND
CC       WHICH DIFFER BOTH IN SEQUENCE AND IN THE NUMBER OF FILAGGRIN
CC       REPEATS.
DR   EMBL; L01089; AAA60177.1; -.
DR   EMBL; L01090; AAA60176.1; -.
DR   HSSP; P02593; 1CDM.
DR   MIM; 135940; -.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003303; Filaggrin.
DR   InterPro; IPR001751; S100_CaBP.
DR   Pfam; PF00036; efhand; 1.
DR   Pfam; PF01023; S_100; 1.
DR   PRINTS; PR00487; FILAGGRIN.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR   PROSITE; PS00303; S100_CABP; 1.
KW   Polymorphism.
FT   PROPEP        1    293       POTENTIAL.
FT   CHAIN       294    467       FILAGGRIN.
FT   PROPEP      468    474       POTENTIAL.
FT   CHAIN       475   >591       FILAGGRIN.
FT   CA_BIND      19     32       SITE I (BY SIMILARITY).
FT   CA_BIND      62     73       SITE II (BY SIMILARITY).
FT   NON_TER     591    591
SQ   SEQUENCE   591 AA;  66366 MW;  381491625C75E369 CRC64;


  Query Match             51.9%;  Score 54;  DB 4;  Length 591;
  Best Local Similarity   64.7%;  Pred. No. 0.45;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||| ||| |  | | |
Db    513 QDTIRGHPGSSRGGRQG 529


RESULT   5
Q9H4U2
ID   Q9H4U2      PRELIMINARY;      PRT;   687 AA.
AC   Q9H4U2;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DJ14N1.1.1 (PROFILAGGRIN 5' END) (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Laird G.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE S-100 FAMILY.
DR   EMBL; AL356504; CAC13172.1; -.
DR   HSSP; P02593; 1CDM.
DR   InterPro; IPR002048; EF-hand.
DR   InterPro; IPR003303; Filaggrin.
DR   InterPro; IPR001751; S100_CaBP.
DR   Pfam; PF01023; S_100; 1.
DR   PRINTS; PR00487; FILAGGRIN.
DR   SMART; SM00054; EFh; 1.
DR   PROSITE; PS00018; EF_HAND; UNKNOWN_1.
FT   NON_TER     687    687
SQ   SEQUENCE   687 AA;  76659 MW;  8000363FBEF07B74 CRC64;


  Query Match             51.9%;  Score 54;  DB 4;  Length 687;
  Best Local Similarity   64.7%;  Pred. No. 0.53;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||| ||| |  | | |
Db    513 QDTIRGHPGSSRGGRQG 529


RESULT   6
Q16824
ID   Q16824      PRELIMINARY;      PRT;   797 AA.
AC   Q16824;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PROFILAGGRIN (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91064347; PubMed=2248957;
RA   Gan S.Q., McBride O.W., Idler W.W., Markova N., Steinert P.M.;
RT   "Organization, structure, and polymorphisms of the human profilaggrin
RT   gene [published erratum appears in Biochemistry 1991 Jun
RT   11;30(23):5814].";
RL   Biochemistry 29:9432-9440(1990).
DR   EMBL; M60502; AAA63248.1; -.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
FT   NON_TER       1      1
SQ   SEQUENCE   797 AA;  85176 MW;  60E6184763BDA86B CRC64;


  Query Match             51.0%;  Score 53;  DB 4;  Length 797;
  Best Local Similarity   64.7%;  Pred. No. 0.9;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||| ||| |  | | |
Db    491 QDTIRGHPGSRRGGRQG 507


RESULT   7
Q9H4U3
ID   Q9H4U3      PRELIMINARY;      PRT;   798 AA.
AC   Q9H4U3;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   DJ14N1.1.2 (PROFILAGGRIN 3' END) (FRAGMENT).
GN   FLG.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Laird G.;
RL   Submitted (OCT-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL356504; CAC13171.1; -.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
FT   NON_TER       1      1
SQ   SEQUENCE   798 AA;  84773 MW;  F923DDA8D1290805 CRC64;


  Query Match             51.0%;  Score 53;  DB 4;  Length 798;
  Best Local Similarity   64.7%;  Pred. No. 0.9;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
```

```
          |||| ||| |  | | |
Db    492 QDTIRGHPGSRRGGRQG 508


RESULT   8
Q01212
ID   Q01212      PRELIMINARY;      PRT;  1084 AA.
AC   Q01212; Q03840;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   FILAGGRIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=PLACENTA;
RX   MEDLINE=91064347; PubMed=2248957;
RA   Gan S.Q., McBride W.O., Idler W.W., Markova N., Steinert P.M.;
RT   "Organization, structure, and polymorphisms of the human profilaggrin
RT   gene.";
RL   Biochemistry 29:9432-9440(1990).
CC   -!- FUNCTION: FILAGGRIN AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND
CC       PROMOTES DISULFID-BOND FORMATION AMONGST THE INTERMEDIATE
CC       FILAMENTS DURING TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC   -!- MISCELLANEOUS: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE,
CC       INSOLUBLE, HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM
CC       COPIES OF 317 AA, WHICH ARE SEPARATED BY A SHORT LINKER SEQUENCE
CC       (PROBABLY FLYQVST). THE PRECURSOR IS DEPOSITED AS KERATOHYALIN
CC       GRANULES. BY MEANS OF DEPHOSPHORYLATION AND PROTEOLYTIC CLEAVAGE
CC       FILAGGRIN IS FORMED.
DR   EMBL; M60503; AAA63243.1; -.
DR   EMBL; M60501; AAA63243.1; JOINED.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
KW   Phosphorylation; Polyprotein; Developmental protein; Keratin;
KW   Intermediate filament.
FT   NON_TER      1      1
SQ   SEQUENCE   1084 AA;  115271 MW;  80C4640B8D5A362D CRC64;


  Query Match             51.0%;  Score 53;  DB 4;  Length 1084;
  Best Local Similarity   64.7%;  Pred. No. 1.2;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||| ||| |  | | |
Db    778 QDTIRGHPGSRRGGRQG 794


RESULT   9
Q89164
ID   Q89164         PRELIMINARY;   PRT;   377 AA.
AC   Q89164;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   ORF7L.
GN   A17L.
OS   Variola virus, and
OS   variola minor virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=10255, 53258;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=Variola virus; STRAIN=GARCIA-1966;
RA   Shchelkunov S.N., Totmenin A.V., Sosnovtsev S.V., Safronov P.F.,
RA   Resenchuk S.M., Blinov V.M., Sandakhchiev L.S.;
RL   Submitted (NOV-1993) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=variola minor virus; STRAIN=GARCIA-1966;
RA   Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
RA   Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA   Esposito J.J., Sosnovtsev S.;
RT   "Analysis of the complete coding sequence of DNA of alastrim variola
RT   minor virus strain Garcia-1966.";
RL   Submitted (MAR-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; X76268; CAA53889.1; -.
DR   EMBL; Y16780; CAB54720.1; -.
DR   InterPro; IPR004251; DUF230.
DR   Pfam; PF03003; DUF230; 1.
SQ   SEQUENCE   377 AA;  43557 MW;  47F10867CB9BE6CE CRC64;


  Query Match             49.0%;  Score 51;  DB 12;  Length 377;
  Best Local Similarity   64.3%;  Pred. No. 0.94;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRPG 17
            ||| |||    |||
Db     85 IHGEPCSSFKFRPG 98


RESULT  10
Q85389
ID   Q85389      PRELIMINARY;      PRT;   377 AA.
AC   Q85389;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   A17L.
GN   A17L.
OS   Variola major virus.
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=12870;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BANGLADESH-1975;
RX   MEDLINE=94088747; PubMed=8264798;
RA   Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA   Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA   Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA   Venter C.J.;
RT   "Potential virulence determinants in terminal regions of variola
RT   smallpox virus genome.";
RL   Nature 366:748-751(1993).
DR   EMBL; L22579; AAA60868.1; -.
DR   InterPro; IPR004251; DUF230.
DR   Pfam; PF03003; DUF230; 1.
SQ   SEQUENCE   377 AA;  43517 MW;  981F36994D6F3093 CRC64;


  Query Match             49.0%;  Score 51;  DB 12;  Length 377;
  Best Local Similarity   64.3%;  Pred. No. 0.94;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      4 IHGHPCSXXGCRPG 17
            ||| |||    |||
Db     85 IHGEPCSSFKFRPG 98


RESULT  11
O93122
ID   O93122      PRELIMINARY;      PRT;   377 AA.
AC   O93122;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   35K MYRISTYLPROTEIN.
GN   MVA127L.
OS   Vaccinia virus (strain Ankara).
```

```
OC   Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC   Orthopoxvirus.
OX   NCBI_TaxID=126794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ANKARA;
RA   Antoine G., Scheiflinger F., Falkner F.G., Dorner F.;
RT   "The complete genomic sequence of the Modified Vaccinia Ankara (MVA)
RT   strain.";
RL   Submitted (MAR-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U94848; AAB96467.1; -.
DR   InterPro; IPR004251; DUF230.
DR   Pfam; PF03003; DUF230; 1.
SQ   SEQUENCE   377 AA;  43428 MW;  EE79C44443A142FA CRC64;


  Query Match             49.0%;  Score 51;  DB 12;  Length 377;
  Best Local Similarity   64.3%;  Pred. No. 0.94;
  Matches    9;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 IHGHPCSXXGCRPG 17
           ||| |||    |||
Db      85 IHGEPCSSFKFRPG 98


RESULT  12
Q9D8P5
ID   Q9D8P5      PRELIMINARY;      PRT;   209 AA.
AC   Q9D8P5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   MAX DIMERIZATION PROTEIN 4.
GN   MAD4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX   MEDLINE=21085660; PubMed=11217851;
RA   Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA   Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA   Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA   Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA   Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA   Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA   Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA   Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA   Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA   Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA   Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA   Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA   Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA   Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA   Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA   Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA   Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA   Hayashizaki Y.;
RT   "Functional annotation of a full-length mouse cDNA collection.";
RL   Nature 409:685-690(2001).
DR   EMBL; AK007824; BAB25287.1; -.
DR   MGD; MGI:104991; Mad4.
DR   InterPro; IPR001092; HLH_dim.
DR   Pfam; PF00010; HLH; 1.
DR   SMART; SM00353; HLH; 1.
SQ   SEQUENCE   209 AA;  23660 MW;  03967C54CE6402D4 CRC64;


  Query Match             48.6%;  Score 50.5;  DB 11;  Length 209;
  Best Local Similarity   56.2%;  Pred. No. 0.64;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPG 17
           |: :|||   || ||
Db     193 DSSYGHPCRRPGC-PG 207


RESULT  13
Q91VN7
ID   Q91VN7      PRELIMINARY;      PRT;   209 AA.
AC   Q91VN7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   MAX DIMERIZATION PROTEIN 4.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=BREAST TUMOR;
RA   Strausberg R.;
RL   Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC011303; AAH11303.1; -.
SQ   SEQUENCE   209 AA;  23614 MW;  02D0BBE70A12F557 CRC64;


  Query Match             48.6%;  Score 50.5;  DB 11;  Length 209;
  Best Local Similarity   56.2%;  Pred. No. 0.64;
  Matches    9;  Conservative    2;  Mismatches    4;  Indels    1;  Gaps    1;

Qy       2 DTIHGHPCSXXGCRPG 17
           |: :|||   || ||
Db     193 DSSYGHPCRRPGC-PG 207


RESULT  14
Q9HH11
ID   Q9HH11      PRELIMINARY;      PRT;   397 AA.
AC   Q9HH11;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   OBG-LIKE PROTEIN.
OS   Thermococcus zilligii.
OC   Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Thermococcus.
OX   NCBI_TaxID=54076;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AN1;
RA   Ronimus R.S., Musgrave D.R.;
RT   "Sequence, transcriptional analysis and phylogeny of a gene from
RT   Thermococcus zilligii encoding a GTP-binding protein with homology to
RT   the essential GTP-binding protein Obg.";
RL   Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY005820; AAF97355.1; -.
DR   InterPro; IPR000765; GTP1_OBG.
DR   InterPro; IPR004095; TGS.
DR   Pfam; PF01018; GTP1_OBG; 3.
DR   Pfam; PF02824; TGS; 1.
DR   PRINTS; PR00326; GTP1OBG.
SQ   SEQUENCE   397 AA;  44424 MW;  7639F66D7BCD73CB CRC64;


  Query Match             48.1%;  Score 50;  DB 1;  Length 397;
  Best Local Similarity   61.5%;  Pred. No. 1.5;
  Matches    8;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       4 IHGHPCSXXGCRP 16
           |  |||   ||||
Db      45 IADHPCKELGCRP 57
```

```
RESULT  15
O75370
ID   O75370      PRELIMINARY;      PRT;   322 AA.
AC   O75370;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EPIDERMAL FILAGGRIN (FRAGMENT).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99101527; PubMed=9886436;
RA   Girbal-Neuhauser E., Durieux J.J., Arnaud M., Dalbon P., Sebbag M.,
RA   Vincent C., Simon M., Senshu T., Masson-Bessiere C.,
RA   Jolivet-Reynaud C., Jolivet M., Serre G.;
RT   "The epitopes targeted by the rheumatoid arthritis-associated
RT   antifilaggrin autoantibodies are posttranslationally generated on
RT   various sites of (pro)filaggrin by deimination of arginine residues.";
RL   J. Immunol. 162:585-594(1999).
DR   EMBL; AF043380; AAC23559.1; -.
DR   InterPro; IPR003303; Filaggrin.
DR   PRINTS; PR00487; FILAGGRIN.
FT   NON_TER       1      1
FT   NON_TER     322    322
SQ   SEQUENCE   322 AA;  34084 MW;  0DC2D0230D8FF9E0 CRC64;


  Query Match             47.1%;  Score 49;  DB 4;  Length 322;
  Best Local Similarity   58.8%;  Pred. No. 1.8;
  Matches   10;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           || | ||| |  | | |
Db      45 QDNIRGHPGSSRGGRQG 61


Search completed: August 26, 2002, 13:35:11
Job time: 366 sec
```

GenCore version 4.5

OM protein - protein search, using sw model

Run on: August 26, 2002, 13:29:01 ; Search time 42.39 Seconds
(without alignments)
47.165 Million cell updates/sec

Title: US-09-747-029A-12
Perfect score: 104
Sequence: 1 QDTIHGHPCSXXGCRPGY 18

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*
1: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1980.DAT:*
2: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1981.DAT:*
3: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1982.DAT:*
4: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1983.DAT:*
5: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1984.DAT:*
6: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1985.DAT:*
7: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1986.DAT:*
8: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1987.DAT:*
9: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1988.DAT:*
10: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1989.DAT:*
11: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1990.DAT:*
12: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1991.DAT:*
13: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1992.DAT:*
14: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1993.DAT:*
15: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1994.DAT:*
16: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1995.DAT:*
17: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1996.DAT:*
18: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1997.DAT:*
19: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1998.DAT:*
20: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA1999.DAT:*
21: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2000.DAT:*
22: /SIDS1/gcgdata/hold-geneseq/geneseqp-embl/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	100	96.2	18	22	AAE07225	IGP1650 peptide fo
2	80	76.9	14	22	AAE07227	IGP1676 peptide fo
3	79	76.0	18	22	AAE07221	IGP1646 peptide fo
4	78	75.0	18	22	AAE07220	IGP1611 peptide fo
5	74	71.2	18	22	AAE07222	IGP1647 peptide fo
6	71	68.3	18	22	AAE07223	IGP1648 peptide fo
7	67	64.4	18	22	AAE07224	IGP1649 peptide fo
8	60	57.7	18	22	AAE07230	IGP1685 peptide fo
9	58	55.8	14	22	AAE07226	IGP1651 peptide fo
10	54	51.9	330	20	AAY22954	Human filagrin seq
11	54	51.9	330	20	AAY22955	Human filagrin seq
12	54	51.9	330	20	AAY22956	Human filagrin seq
13	54	51.9	330	20	AAY22957	Human filagrin seq
14	51.5	49.5	149	22	ABG11403	Novel human diagno
15	51.5	49.5	810	18	AAW37500	Human nel-related
16	51	49.0	72	22	AAU56399	Propionibacterium
17	51	49.0	254	22	AAU40819	Propionibacterium
18	47	45.2	724	13	AAR27648	Human calcium chan
19	47	45.2	2251	16	AAR71009	Human neuronal cal
20	47	45.2	2251	21	AAB10581	Human calcium chan
21	47	45.2	2270	16	AAR71010	Human neuronal cal
22	47	45.2	2270	16	AAR69604	Calcium channel al
23	47	45.2	2270	21	AAB10582	Human calcium chan
24	46	44.2	24	22	ABB29263	Peptide #1914 enco
25	46	44.2	24	22	ABB34433	Peptide #1939 enco
26	46	44.2	24	22	ABB19843	Protein #1842 enco
27	46	44.2	24	22	AAM55219	Human brain expres
28	46	44.2	24	22	AAM67615	Human bone marrow
29	46	44.2	24	22	AAM15421	Peptide #1855 enco
30	46	44.2	24	22	AAM03183	Peptide #1865 enco
31	46	44.2	122	21	AAB24464	Human secreted pro
32	46	44.2	349	22	ABG02912	Novel human diagno
33	46	44.2	397	22	AAB96502	Putative P. abyssi
34	45.5	43.8	240	19	AAW64219	Human secreted pro
35	45.5	43.8	240	22	AAB90729	Human CB107_1 pro
36	45.5	43.8	274	22	ABB58794	Drosophila melanog
37	45.5	43.8	816	18	AAW37501	Human nel-related
38	44.5	42.8	1480	13	AAR25079	Drosophila SLIT pr
39	44.5	42.8	4601	22	ABB59371	Drosophila melanog
40	44	42.3	69	22	AAU50528	Propionibacterium
41	44	42.3	529	22	ABB70849	Drosophila melanog
42	44	42.3	945	20	AAY35612	C. pneumoniae prot
43	44	42.3	1224	21	AAB18258	Plasmodium falcipa
44	43.5	41.8	80	22	AAU39827	Propionibacterium
45	43.5	41.8	159	22	AAY72740	HMEIR04 clone huma

ALIGNMENTS

RESULT 1
AAE07225
ID AAE07225 standard; peptide; 18 AA.
XX
AC AAE07225;
XX
DT 06-NOV-2001 (first entry)
XX
DE IGP1650 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW Synthetic peptide; cyclic; IGP1650; autoimmune antibody;
KW rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Modified-site 1..18
FT /note= "Biotinylated residues"
FT Disulfide-bond 9..14
FT Modified-site 11
FT /note= "Citrulline"
FT Modified-site 12
FT /note= "Citrulline"
XX
PN WO200146222-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000WO-EP13037.
XX
PR 21-DEC-1999; 99EP-0870280.
PR 08-SEP-2000; 2000EP-0870195.
XX

```
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1650.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             96.2%;  Score 100;  DB 22;  Length 18;
  Best Local Similarity   100.0%;  Pred. No. 3.9e-09;
  Matches   18;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||||||||||||||||||
Db       1 qdtihghpcsxxgcrpgy 18


RESULT   2
AAE07227
ID   AAE07227 standard; peptide; 14 AA.
XX
AC   AAE07227;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1676 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1676; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..14
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..14
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1676.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   14 AA;


  Query Match             76.9%;  Score 80;  DB 22;  Length 14;
  Best Local Similarity   100.0%;  Pred. No. 4.4e-06;
  Matches   14;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       5 HGHPCSXXGCRPGY 18
           ||||||||||||||
Db       1 hghpcsxxgcrpgy 14


RESULT   3
AAE07221
ID   AAE07221 standard; peptide; 18 AA.
XX
AC   AAE07221;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1646 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1646; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1646.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;
```

```
  Query Match             76.0%;  Score 79;  DB 22;  Length 18;
  Best Local Similarity   83.3%;  Pred. No. 8e-06;
  Matches   15;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||||||||||  || | ||
Db       1 qdtihghpcssxghrcgy 18
```

```
RESULT   4
AAE07220
ID   AAE07220 standard; peptide; 18 AA.
XX
AC   AAE07220;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1611 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1611; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1611.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;
```

```
  Query Match             75.0%;  Score 78;  DB 22;  Length 18;
  Best Local Similarity   88.9%;  Pred. No. 1.2e-05;
  Matches   16;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||||||||||||| | ||
Db       1 qdtihghpcsxxghrcgy 18
```

```
RESULT   5
AAE07222
ID   AAE07222 standard; peptide; 18 AA.
XX
AC   AAE07222;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1647 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1647; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1647.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             71.2%;  Score 74;  DB 22;  Length 18;
  Best Local Similarity   83.3%;  Pred. No. 4.9e-05;
  Matches   15;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPGY 18
           ||||||||||||| : ||
Db       1 qdtihghpcsxxghqcgy 18


RESULT   6
AAE07223
ID   AAE07223 standard; peptide; 18 AA.
XX
AC   AAE07223;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1648 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1648; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1648.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             68.3%;  Score 71;  DB 22;  Length 18;
  Best Local Similarity   88.2%;  Pred. No. 0.00015;
  Matches   15;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           |||||||||||||| | |
Db       1 qdtihghpcsxxghrcg 17


RESULT   7
AAE07224
ID   AAE07224 standard; peptide; 18 AA.
XX
AC   AAE07224;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1649 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1649; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
```

```
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1649.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             64.4%;  Score 67;  DB 22;  Length 18;
  Best Local Similarity   82.4%;  Pred. No. 0.00063;
  Matches   14;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          |||||||||||||  : |
Db      1 qdtihghpcsxxghqcg 17


RESULT   8
AAE07230
ID   AAE07230 standard; peptide; 18 AA.
XX
AC   AAE07230;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1685 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1685; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..18
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..14
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1685.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   18 AA;


  Query Match             57.7%;  Score 60;  DB 22;  Length 18;
  Best Local Similarity   70.6%;  Pred. No. 0.008;
  Matches   12;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy      1 QDTIHGHPCSXXGCRPG 17
          ||||  |  |  ||||||
Db      1 qdtivgwgcdsxgcrpg 17


RESULT   9
AAE07226
ID   AAE07226 standard; peptide; 14 AA.
XX
AC   AAE07226;
XX
DT   06-NOV-2001  (first entry)
XX
DE   IGP1651 peptide for diagnosis and treatment of rheumatoid arthritis.
XX
KW   Synthetic peptide; cyclic; IGP1651; autoimmune antibody;
KW   rheumatoid arthritis; therapy; autoimmune disease; antirheumatic;
KW   systemic hyporesponsiveness; immunosuppressive; antiarthritic.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Modified-site   1..14
FT                   /note= "Biotinylated residues"
FT   Disulfide-bond  9..16
FT   Modified-site   11
FT                   /note= "Citrulline"
FT   Modified-site   12
FT                   /note= "Citrulline"
XX
PN   WO200146222-A2.
XX
PD   28-JUN-2001.
XX
PF   20-DEC-2000; 2000WO-EP13037.
XX
PR   21-DEC-1999;   99EP-0870280.
PR   08-SEP-2000; 2000EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Union A,  Moereels H,  Meheus L;
XX
```

```
DR   WPI; 2001-496657/54.
XX
PT   New peptides, useful for diagnosing and treating rheumatoid arthritis,
PT   comprises citrulline residue between 2 cysteine residues and is
PT   specifically recognized by autoimmune antibodies from patients
PT   suffering from rheumatoid arthritis -
XX
PS   Claim 9; Page 42; 53pp; English.
XX
CC   The present sequence is a cyclic synthetic biotinylated peptide, IGP1651.
CC   The peptide comprises a citrulline residue between 2 cysteine residues
CC   and is specifically recognised by autoimmune antibodies from patients
CC   suffering from rheumatoid arthritis. The peptide comprises amino acids
CC   involved in side chain interactions which is essential for the formation
CC   of three-dimensional structure of the peptide. The peptide of the
CC   invention is useful as a medicament to treat autoimmune diseases,
CC   preferably rheumatoid arthritis. It is also useful for treating
CC   autoimmune diseases by increasing the size of antigen-immune complexes to
CC   improve clearance of the formed immune complexes and for the preparation
CC   of a medicament for oral or nasal administration to treat autoimmune
CC   diseases by inducing a state of systemic hyporesponsiveness or tolerance
CC   to the peptide.
XX
SQ   Sequence   14 AA;


  Query Match             55.8%;  Score 58;  DB 22;  Length 14;
  Best Local Similarity   85.7%;  Pred. No. 0.013;
  Matches   12;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       5 HGHPCSXXGCRPGY 18
           |||||||||  | ||
Db       1 hghpcsxxghrcgy 14


RESULT  10
AAY22954
ID   AAY22954 standard; peptide; 330 AA.
XX
AC   AAY22954;
XX
DT   20-AUG-1999  (first entry)
XX
DE   Human filagrin sequence of clone HB2641.
XX
KW   Filaggrin; intermediate filament protein; antibody; rheumatoid arthritis;
KW   antigen; immunotoxin; autoantigen; autoantibody; autoimmune disease;
KW   systemic lupus erythematosus; discoid lupus erythematosus; scleroderma;
KW   dermatomyositis; Sjogrens syndrome.
XX
OS   Homo sapiens.
XX
PN   WO9928344-A2.
XX
PD   10-JUN-1999.
XX
PF   30-NOV-1998;   98WO-EP07714.
XX
PR   09-APR-1998;   98EP-0870078.
PR   28-NOV-1997;   97EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Meheus L,  Raymackers J,  Union A;
XX
DR   WPI; 1999-385357/32.
XX
PT   New peptide derived from intermediate filament proteins
XX
PS   Example 1; Fig 2; 73pp; English.
XX
CC   AAY22954-57 represent amino acid sequences of human filagrin clones. The
CC   specification describes peptides derived from any variant of natural
CC   filaggrin or any variant of intermediate filament proteins. These
CC   peptides contain at least one citrulline residue which is crucial
CC   for reacting with antibodies that are present in sera from patients
CC   with rheumatoid arthritis. The peptides constitute immunogenic
CC   determinants of antibodies present in patients with rheumatoid
CC   arthritis. The peptides, antibodies, immunotoxins and intermediate
CC   filament proteins can be used for the preparation of a therapeutic or
CC   of a diagnostic for rheumatoid arthritis. The peptides can also be
CC   used for identifying compounds which modulate the interaction between
CC   an autoantigen and a rheumatoid arthritis specific autoantibody. The
CC   products can also be used for the diagnosis and treatment of other
CC   autoimmune diseases e.g. systemic lupus erythematosus, discoid lupus
CC   erythematosus, scleroderma, dermatomyositis, or Sjogrens syndrome.
XX
SQ   Sequence   330 AA;


  Query Match             51.9%;  Score 54;  DB 20;  Length 330;
  Best Local Similarity   64.7%;  Pred. No. 1.2;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           |||||||  |  | | |
Db      49 qdtihghrgsssggrqg 65


RESULT  11
AAY22955
ID   AAY22955 standard; peptide; 330 AA.
XX
AC   AAY22955;
XX
DT   20-AUG-1999  (first entry)
XX
DE   Human filagrin sequence of clone HB2642.
XX
KW   Filaggrin; intermediate filament protein; antibody; rheumatoid arthritis;
KW   antigen; immunotoxin; autoantigen; autoantibody; autoimmune disease;
KW   systemic lupus erythematosus; discoid lupus erythematosus; scleroderma;
KW   dermatomyositis; Sjogrens syndrome.
XX
OS   Homo sapiens.
XX
PN   WO9928344-A2.
XX
PD   10-JUN-1999.
XX
PF   30-NOV-1998;   98WO-EP07714.
XX
PR   09-APR-1998;   98EP-0870078.
PR   28-NOV-1997;   97EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Meheus L,  Raymackers J,  Union A;
XX
DR   WPI; 1999-385357/32.
XX
PT   New peptide derived from intermediate filament proteins
XX
PS   Example 1; Fig 2; 73pp; English.
XX
CC   AAY22954-57 represent amino acid sequences of human filagrin clones. The
CC   specification describes peptides derived from any variant of natural
CC   filaggrin or any variant of intermediate filament proteins. These
CC   peptides contain at least one citrulline residue which is crucial
CC   for reacting with antibodies that are present in sera from patients
CC   with rheumatoid arthritis. The peptides constitute immunogenic
CC   determinants of antibodies present in patients with rheumatoid
CC   arthritis. The peptides, antibodies, immunotoxins and intermediate
CC   filament proteins can be used for the preparation of a therapeutic or
CC   of a diagnostic for rheumatoid arthritis. The peptides can also be
CC   used for identifying compounds which modulate the interaction between
```

```
CC   an autoantigen and a rheumatoid arthritis specific autoantibody. The
CC   products can also be used for the diagnosis and treatment of other
CC   autoimmune diseases e.g. systemic lupus erythematosus, discoid lupus
CC   erythematosus, scleroderma, dermatomyositis, or Sjogrens syndrome.
XX
SQ   Sequence   330 AA;


  Query Match             51.9%;  Score 54;  DB 20;  Length 330;
  Best Local Similarity   64.7%;  Pred. No. 1.2;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           |||||||  |  | | |
Db      49 qdtihghrgsssggrqg 65


RESULT  12
AAY22956
ID   AAY22956 standard; peptide; 330 AA.
XX
AC   AAY22956;
XX
DT   20-AUG-1999  (first entry)
XX
DE   Human filagrin sequence of clone HB2650.
XX
KW   Filaggrin; intermediate filament protein; antibody; rheumatoid arthritis;
KW   antigen; immunotoxin; autoantigen; autoantibody; autoimmune disease;
KW   systemic lupus erythematosus; discoid lupus erythematosus; scleroderma;
KW   dermatomyositis; Sjogrens syndrome.
XX
OS   Homo sapiens.
XX
PN   WO9928344-A2.
XX
PD   10-JUN-1999.
XX
PF   30-NOV-1998;   98WO-EP07714.
XX
PR   09-APR-1998;   98EP-0870078.
PR   28-NOV-1997;   97EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Meheus L,  Raymackers J,  Union A;
XX
DR   WPI; 1999-385357/32.
XX
PT   New peptide derived from intermediate filament proteins
XX
PS   Example 1; Fig 2; 73pp; English.
XX
CC   AAY22954-57 represent amino acid sequences of human filagrin clones. The
CC   specification describes peptides derived from any variant of natural
CC   filaggrin or any variant of intermediate filament proteins. These
CC   peptides contain at least one citrulline residue which is crucial
CC   for reacting with antibodies that are present in sera from patients
CC   with rheumatoid arthritis. The peptides constitute immunogenic
CC   determinants of antibodies present in patients with rheumatoid
CC   arthritis. The peptides, antibodies, immunotoxins and intermediate
CC   filament proteins can be used for the preparation of a therapeutic or
CC   of a diagnostic for rheumatoid arthritis. The peptides can also be
CC   used for identifying compounds which modulate the interaction between
CC   an autoantigen and a rheumatoid arthritis specific autoantibody. The
CC   products can also be used for the diagnosis and treatment of other
CC   autoimmune diseases e.g. systemic lupus erythematosus, discoid lupus
CC   erythematosus, scleroderma, dermatomyositis, or Sjogrens syndrome.
XX
SQ   Sequence   330 AA;

  Query Match             51.9%;  Score 54;  DB 20;  Length 330;
  Best Local Similarity   64.7%;  Pred. No. 1.2;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           |||||||  |  | | |
Db      49 qdtihghrgsssggrqg 65


RESULT  13
AAY22957
ID   AAY22957 standard; peptide; 330 AA.
XX
AC   AAY22957;
XX
DT   20-AUG-1999  (first entry)
XX
DE   Human filagrin sequence of clone HB2648.
XX
KW   Filaggrin; intermediate filament protein; antibody; rheumatoid arthritis;
KW   antigen; immunotoxin; autoantigen; autoantibody; autoimmune disease;
KW   systemic lupus erythematosus; discoid lupus erythematosus; scleroderma;
KW   dermatomyositis; Sjogrens syndrome.
XX
OS   Homo sapiens.
XX
PN   WO9928344-A2.
XX
PD   10-JUN-1999.
XX
PF   30-NOV-1998;   98WO-EP07714.
XX
PR   09-APR-1998;   98EP-0870078.
PR   28-NOV-1997;   97EP-0870195.
XX
PA   (INNO-) INNOGENETICS NV.
XX
PI   Meheus L,  Raymackers J,  Union A;
XX
DR   WPI; 1999-385357/32.
XX
PT   New peptide derived from intermediate filament proteins
XX
PS   Example 1; Fig 2; 73pp; English.
XX
CC   AAY22954-57 represent amino acid sequences of human filagrin clones. The
CC   specification describes peptides derived from any variant of natural
CC   filaggrin or any variant of intermediate filament proteins. These
CC   peptides contain at least one citrulline residue which is crucial
CC   for reacting with antibodies that are present in sera from patients
CC   with rheumatoid arthritis. The peptides constitute immunogenic
CC   determinants of antibodies present in patients with rheumatoid
CC   arthritis. The peptides, antibodies, immunotoxins and intermediate
CC   filament proteins can be used for the preparation of a therapeutic or
CC   of a diagnostic for rheumatoid arthritis. The peptides can also be
CC   used for identifying compounds which modulate the interaction between
CC   an autoantigen and a rheumatoid arthritis specific autoantibody. The
CC   products can also be used for the diagnosis and treatment of other
CC   autoimmune diseases e.g. systemic lupus erythematosus, discoid lupus
CC   erythematosus, scleroderma, dermatomyositis, or Sjogrens syndrome.
XX
SQ   Sequence   330 AA;


  Query Match             51.9%;  Score 54;  DB 20;  Length 330;
  Best Local Similarity   64.7%;  Pred. No. 1.2;
  Matches   11;  Conservative    0;  Mismatches    6;  Indels    0;  Gaps    0;

Qy       1 QDTIHGHPCSXXGCRPG 17
           |||| |||  |  | | |
Db      49 qdtirghpgssrggrqg 65


RESULT  14
```

```
ABG11403
ID   ABG11403 standard; Protein; 149 AA.
XX
AC   ABG11403;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #11394.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS75590.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 20; SEQ ID No 41762; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   149 AA;


  Query Match             49.5%;  Score 51.5;  DB 22;  Length 149;
  Best Local Similarity   47.1%;  Pred. No. 1.4;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    3;  Gaps    1;

Qy      2 DTIHGHPCSXXGCRPGY 18
          :|: || |:   |:|||
Db     37 ntvqghsct---ckpgy 50


RESULT  15
AAW37500
ID   AAW37500 standard; Protein; 810 AA.
XX
AC   AAW37500;
XX
DT   20-APR-1998  (first entry)
XX
DE   Human nel-related protein type 1.
XX
KW   Human; foetal brain cDNA library; GDP dissociation stimulating protein;
KW   brain specific nucleosome assembly protein; diagnosis; therapy;
KW   skeletal muscle specific ubiquitin conjugating enzyme; TMP-2; NPIK;
KW   nel-related protein type 1; nel-related type 2; hereditary disease;
KW   cancer.
XX
OS   Homo sapiens.
XX
PN   EP796913-A2.
XX
PD   24-SEP-1997.
XX
PF   19-MAR-1997;   97EP-0104842.
XX
PR   05-MAR-1997;   97JP-0069163.
PR   19-MAR-1996;   96JP-0063410.
XX
PA   (SAKA ) OTSUKA PHARM CO LTD.
XX
PI   Fujiwara T,  Horie M,  Watanabe T;
XX
DR   WPI; 1997-459830/43.
DR   N-PSDB; AAV01880, AAV01881.
XX
PT   Novel human genes, e.g. brain-specific nucleosome assembly protein -
PT   useful for diagnosis or therapy of hereditary disease and cancer
XX
PS   Claim 19; Page 94-97; 123pp; English.
XX
CC   The present sequence represents a nel-related protein type 1 isolated
CC   from a human foetal brain cDNA library. The nucleotide or amino acid
CC   sequences are useful for in-vitro diagnosis of hereditary diseases and
CC   cancer and for preparation of pharmaceuticals.
XX
SQ   Sequence   810 AA;


  Query Match             49.5%;  Score 51.5;  DB 18;  Length 810;
  Best Local Similarity   47.1%;  Pred. No. 7.3;
  Matches    8;  Conservative    4;  Mismatches    2;  Indels    3;  Gaps    1;

Qy      2 DTIHGHPCSXXGCRPGY 18
          :|: || |:   |:|||
Db    495 ntvqghsct---ckpgy 508


Search completed: August 26, 2002, 13:29:01
Job time: 66 sec
```

```
=> fil reg
FILE 'REGISTRY' ENTERED AT 15:24:23 ON 26 AUG 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 American Chemical Society (ACS)

STRUCTURE FILE UPDATES:   25 AUG 2002  HIGHEST RN 444874-82-2
DICTIONARY FILE UPDATES:  25 AUG 2002  HIGHEST RN 444874-82-2

TSCA INFORMATION NOW CURRENT THROUGH MAY 20, 2002

  Please note that search-term pricing does apply when
  conducting SmartSELECT searches.

Crossover limits have been increased.  See HELP CROSSOVER for details.

Calculated physical property data is now available.  See HELP PROPERTIES
for more information.  See STNote 27, Searching Properties in the CAS
Registry File, for complete details:
http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf

=> d sta que l6
L1            1 SEA FILE=REGISTRY ABB=ON  PLU=ON  (QDTIHGHPCS'AAA-AAA'GCRPGY)/S
                QEP
L2            1 SEA FILE=REGISTRY ABB=ON  PLU=ON  QDTIHGHPCS..GCRPGY/SQSP
L3            1 SEA FILE=REGISTRY ABB=ON  PLU=ON  (L1 OR L2)
L5            1 SEA FILE=REGISTRY ABB=ON  PLU=ON  [STDENQHKR][STDENQHKR][STDENQ
                HKR][ILVAM][HKRDESTYFW]G[HKRDESTYFW][PG]C[STDG]..GC[RKHDESTQNYF
                W][PG]G[YHKRDESTQNFW]/SQSP
L6            1 SEA FILE=REGISTRY ABB=ON  PLU=ON  (L3 OR L5)


=> d his

     (FILE 'HOME' ENTERED AT 15:05:52 ON 26 AUG 2002)
                SET COST OFF

     FILE 'REGISTRY' ENTERED AT 15:06:42 ON 26 AUG 2002
                E QDTIHGHPCS'CIT''CIT'GCRPGY/SQEP
L1            1 S E1
L2            1 S QDTIHGHPCS..GCRPGY/SQSP
L3            1 S L1,L2
L4            0 S [STDENQHKR][STDENQHKR][STDENQHKR][ILVAM][HKRDESTYFW]G[HKRDEST
L5            1 S [STDENQHKR][STDENQHKR][STDENQHKR][ILVAM][HKRDESTYFW]G[HKRDEST
L6            1 S L3,L5
                SAV L6 DIBRINO747/A

     FILE 'HCAOLD' ENTERED AT 15:12:44 ON 26 AUG 2002
L7            0 S L6

     FILE 'USPATFULL, USPAT2' ENTERED AT 15:12:47 ON 26 AUG 2002
L8            0 S L6

     FILE 'HCAPLUS' ENTERED AT 15:12:52 ON 26 AUG 2002
L9            1 S L6
                SEL RN

     FILE 'REGISTRY' ENTERED AT 15:13:12 ON 26 AUG 2002
L10          19 S E1-E19
L11          18 S L10 NOT L6
L12           3 S L11 NOT SQL/FA
L13          15 S L11 NOT L12
L14          10 S L13 AND SQL<=18
```

Jan Delaval
Reference Librarian
Biotechnology & Chemical Library
CM1 1E07 – 703-308-4498
jan.delaval@uspto.gov

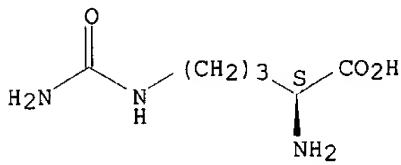
```
     FILE 'HCAPLUS' ENTERED AT 15:17:39 ON 26 AUG 2002
                E UNION A/AU
L15            11 S E3,E4
                E MOEREELS H/AU
L16            55 S E3,E4,E6-E8
                E MEHEUS L/AU
L17            25 S E3-E5
                E INNOGENET/PA,CS
L18           166 S E4-E40
L19             1 S L15-L18 AND L9
L20           240 S L15-L18 NOT L19
L21             3 S L20 AND ?CITRUL?

     FILE 'REGISTRY' ENTERED AT 15:20:39 ON 26 AUG 2002
L22             1 S L12 AND ORNITH?
                E D-CITRULLINE/CN
L23             1 S E3
                E DL-CITRULLINE/CN
L24             1 S E3

     FILE 'HCAPLUS' ENTERED AT 15:21:06 ON 26 AUG 2002
L25             3 S L22 AND L15-L18
L26             4 S L21,L25
L27             4 S L26 AND ?CITRUL?
L28             3 S L27 NOT L19
                SEL RN

     FILE 'REGISTRY' ENTERED AT 15:21:46 ON 26 AUG 2002
L29            10 S E1-E10
L30             1 S L29 AND N5
L31             1 S L29 AND AMINOCARBONY?
L32             1 S L29 AND ORNITH?
L33             1 S L30-L32

     FILE 'REGISTRY' ENTERED AT 15:24:23 ON 26 AUG 2002

=> d sqide can 16

L6   ANSWER 1 OF 1  REGISTRY  COPYRIGHT 2002 ACS
RN   347873-68-1  REGISTRY
CN   Peptide, (Gln-Asp-Thr-Ile-His-Gly-His-Pro-Cys-Ser-Xaa-Xaa-Gly-Cys-Arg-Pro-
     Gly-Tyr) (9CI)  (CA INDEX NAME)
OTHER NAMES:
CN   20: PN: WO0146222 SEQID: 6 claimed protein
CN   6: PN: WO0146222 PAGE: 27 claimed sequence
FS   PROTEIN SEQUENCE
SQL  18
NTE
```

type	location		description
uncommon	Aaa-11	-	-
uncommon	Aaa-12	-	-

PATENT ANNOTATIONS (PNTE):

Sequence Source	Patent Reference
Not Given	WO2001046222 claimed PAGE 27

```
---------+-----------
         |WO2001046222
         |claimed
         |SEQID 6


SEQ      1 QDTIHGHPCS XXGCRPGY
           ========== ========
HITS AT:   1-18
MF   Unspecified
CI   MAN
SR   CA
LC   STN Files:   CA, CAPLUS, TOXCENTER
               1 REFERENCES IN FILE CA (1967 TO DATE)
               1 REFERENCES IN FILE CAPLUS (1967 TO DATE)

REFERENCE   1:  135:91514


=> d l22 ide can

L22  ANSWER 1 OF 1  REGISTRY  COPYRIGHT 2002 ACS
RN   372-75-8  REGISTRY
CN   L-Ornithine, N5-(aminocarbonyl)- (9CI)  (CA INDEX NAME)
OTHER CA INDEX NAMES:
CN   Ornithine, N5-carbamoyl-, L- (8CI)
OTHER NAMES:
CN   .alpha.-Amino-.delta.-ureidovaleric acid
CN   .delta.-Ureidonorvaline
CN   Citrulline
CN   L-Citrulline
CN   N.delta.-Carbamylornithine
CN   N5-Carbamoyl-L-ornithine
FS   STEREOSEARCH
MF   C6 H13 N3 O3
CI   COM
LC   STN Files:   ADISNEWS, AGRICOLA, ANABSTR, BEILSTEIN*, BIOBUSINESS, BIOSIS,
       BIOTECHNO, CA, CANCERLIT, CAOLD, CAPLUS, CASREACT, CEN, CHEMCATS,
       CHEMINFORMRX, CHEMLIST, CIN, CSCHEM, DDFU, DRUGU, EMBASE, HODOC*,
       IFICDB, IFIPAT, IFIUDB, IPA, MEDLINE, MRCK*, MSDS-OHS, NAPRALERT,
       NIOSHTIC, PIRA, PROMT, SPECINFO, TOXCENTER, USPAT2, USPATFULL, VETU
         (*File contains numerically searchable property data)
     Other Sources:   EINECS**, NDSL**, TSCA**
         (**Enter CHEMLIST File for up-to-date regulatory information)

Absolute stereochemistry.
```




```
**PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT**

            2848 REFERENCES IN FILE CA (1967 TO DATE)
              48 REFERENCES TO NON-SPECIFIC DERIVATIVES IN FILE CA
            2851 REFERENCES IN FILE CAPLUS (1967 TO DATE)
```

69 REFERENCES IN FILE CAOLD (PRIOR TO 1967)

REFERENCE 1: 137:124407

REFERENCE 2: 137:122202

REFERENCE 3: 137:98762

REFERENCE 4: 137:92082

REFERENCE 5: 137:91230

REFERENCE 6: 137:83616

REFERENCE 7: 137:78493

REFERENCE 8: 137:76169

REFERENCE 9: 137:75434

REFERENCE 10: 137:62634

=> fil hcaplus
FILE 'HCAPLUS' ENTERED AT 15:25:02 ON 26 AUG 2002
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2002 AMERICAN CHEMICAL SOCIETY (ACS)

FILE COVERS 1907 - 26 Aug 2002 VOL 137 ISS 9
FILE LAST UPDATED: 25 Aug 2002 (20020825/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

CAS roles have been modified effective December 16, 2001. Please check your SDI profiles to see if they need to be revised. For information on CAS roles, enter HELP ROLES at an arrow prompt or use the CAS Roles thesaurus (/RL field) in this file.

=> d all hitstr l9

L9 ANSWER 1 OF 1 HCAPLUS COPYRIGHT 2002 ACS
AN 2001:472747 HCAPLUS
DN 135:91514
TI Peptides designed for the diagnosis and treatment of rheumatoid arthritis
IN Union, Ann; Moereels, Henri; Meheus, Lydie
PA Innogenetics N.V., Belg.
SO PCT Int. Appl., 53 pp.
CODEN: PIXXD2
DT Patent
LA English

IC ICM C07K007-08
CC 15-2 (Immunochemistry)
Section cross-reference(s): 9, 63
FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 2001046222	A2	20010628	WO 2000-EP13037	20001220
	WO 2001046222	A3	20020117		

W: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM
RW: GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG

PRAI EP 1999-870280 A 19991221
EP 2000-870195 A 20000908

AB The present invention relates to peptides that mimic the immunogenic determinants of self-proteins recognized by autoimmune antibodies in a biol. sample from patients suffering from rheumatoid arthritis (RA). More particularly, the present invention relates to citrulline-contg. peptides, which react with the majority of the latter antibodies. Furthermore, the present invention relates to diagnostic tools for a more convenient and sensitive diagnosis of RA and to therapeutical methods to treat RA.

ST autoimmune disease rheumatoid arthritis citrulline peptide

IT Diagnosis
(agents; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Antibodies
RL: BSU (Biological study, unclassified); MFM (Metabolic formation); THU (Therapeutic use); BIOL (Biological study); FORM (Formation, nonpreparative); USES (Uses)
(anti-idiotypic; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Antibodies
RL: BPR (Biological process); BSU (Biological study, unclassified); BIOL (Biological study); PROC (Process)
(autoantibodies; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Antigens
RL: BSU (Biological study, unclassified); BIOL (Biological study)
(autoantigens; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Autoimmune disease
Blood serum
Immune tolerance
Protein sequences
Rheumatoid arthritis
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Filaggrin
RL: BSU (Biological study, unclassified); BIOL (Biological study)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Antibodies
RL: BSU (Biological study, unclassified); MFM (Metabolic formation); THU (Therapeutic use); BIOL (Biological study); FORM (Formation, nonpreparative); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Peptides, biological studies

RL: BSU (Biological study, unclassified); PRP (Properties); SPN (Synthetic preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Antigens
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Peptides, biological studies
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(cyclic; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Test kits
(diagnostic; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Diagnosis
(immunodiagnosis; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Drug delivery systems
(immunotoxins; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Antibodies
RL: BSU (Biological study, unclassified); MFM (Metabolic formation); THU (Therapeutic use); BIOL (Biological study); FORM (Formation, nonpreparative); USES (Uses)
(monoclonal; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Drug delivery systems
(nasal; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Diagnosis
(serodiagnosis; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT Membranes, nonbiological
(strip solid support; citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT 75536-80-0, Peptidylarginine deiminase
RL: BSU (Biological study, unclassified); BUU (Biological use, unclassified); BIOL (Biological study); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT 347871-56-1P 347871-73-2P 347871-78-7P 347872-77-9P 347873-22-7P
347873-68-1P 347873-98-7P 347874-24-2P 347874-53-7P
347874-78-6P 347875-05-2P 347875-19-8P
RL: BSU (Biological study, unclassified); PRP (Properties); SPN (Synthetic preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT 58-85-5, Biotin 372-75-8, Citrulline
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)

IT 347875-37-0 347875-54-1 347875-70-1 347875-88-1
RL: PRP (Properties)
(unclaimed sequence; peptides designed for the diagnosis and treatment of rheumatoid arthritis)

IT **347873-68-1P**
RL: BSU (Biological study, unclassified); PRP (Properties); SPN (Synthetic

preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(citrulline-contg. peptides for diagnosis and treatment of rheumatoid arthritis)
RN 347873-68-1 HCAPLUS
CN Peptide, (Gln-Asp-Thr-Ile-His-Gly-His-Pro-Cys-Ser-Xaa-Xaa-Gly-Cys-Arg-Pro-Gly-Tyr) (9CI) (CA INDEX NAME)

*** STRUCTURE DIAGRAM IS NOT AVAILABLE ***

=> d all tot l28

L28 ANSWER 1 OF 3 HCAPLUS COPYRIGHT 2002 ACS
AN 2002:448386 HCAPLUS
TI Identification of **citrullinated** rheumatoid arthritis-specific epitopes in natural filaggrin relevant for antifilaggrin autoantibody detection by line immunoassay
AU **Union, Ann; Meheus, Lydie**; Humbel, Rene Louis; Conrad, Karsten; Steiner, Guenter; **Moereels, Henri**; Pottel, Hans; Serre, Guy; De Keyser, Filip
CS **Innogenetics NV, Ghent, 9052, Belg.**
SO Arthritis & Rheumatism (2002), 46(5), 1185-1195
CODEN: ARHEAW; ISSN: 0004-3591
PB John Wiley & Sons, Inc.
DT Journal
LA English
CC 15 (Immunochemistry)
AB To identify immunodominant epitopes in natural filaggrin that are reactive with antifilaggrin autoantibodies (AFA) in the sera of patients with rheumatoid arthritis (RA) and to explore their use in a diagnostic assay format. Based on the results of epitope mapping of human natural filaggrin as well as mol. modeling and computational chem., synthetic peptides together with recombinant **citrullinated** filaggrin were evaluated by a line immunoassay (LIA) for AFA detection. Diagnostic performance was assessed using 336 RA and 253 disease control sera and was compared with that of ref. methods. Several immunoreactive epitopes were identified in natural filaggrin, all of which contained at least 1 **citrulline** residue. Three antigenic substrates, including 2 synthetic peptides and recombinant **citrullinated** filaggrin showing maximal reactivity on LIA, were finally selected. Using the 3-antigen LIA3, overall sensitivity, specificity, and pos. predictive value for RA were 65.2%, 98.0%, and 89.1%, resp., compared with 61.9%, 98.8%, and 92.8% using the 2-antigen LIA2 (without recombinant protein). Thirty-seven percent of the rheumatoid factor (RF)-neg. RA samples (30 of 81) were AFA-pos. by LIA2, and 52 of 54 RF-pos. control samples had no AFA detected on LIA2. Higher specificity and sensitivity were obtained by LIA2 vs. anti-RA33 immunoblot, whereas good agreement was obsd. with antikeratin antibody testing. LIA performed significantly better than AFA immunoblotting using natural filaggrin, at a specificity level of 99% (P = 0.0047). **Citrullinated** residues are present in immunoreactive epitopes of natural human filaggrin. AFA can be readily detected by **citrullinated** peptides in an LIA-based test, resulting in high specificity and pos. predictive value for RA. The LIA could serve as a user-friendly alternative to existing immunofluorescence tests and AFA immunoblot techniques. Given its complementarity to RF, this test can be a valuable tool in the differential diagnosis of arthritis.
RE.CNT 40 THERE ARE 40 CITED REFERENCES AVAILABLE FOR THIS RECORD
RE
(1) Aho, K; J Rheumatol 1993, V34, P1278
(2) Aho, K; Scand J Rheumatol 1999, V28, P113 MEDLINE
(3) Arnett, F; Arthritis Rheum 1988, V31, P315 MEDLINE
(4) Asaga, H; Biochem Biophys Res Commun 1998, V243, P641 HCAPLUS

(5) Baeten, D; Arthritis Rheum 2001, V44, P2255 HCAPLUS
(6) Berthelot, J; Ann Rheum Dis 1997, V56, P123 MEDLINE
(7) Brahms, H; J Biol Chem 2000, V275, P17122 HCAPLUS
(8) Cordonnier, C; Br J Rheumatol 1996, V35, P620 MEDLINE
(9) Firestein, G; Am J Pathol 1996, V149, P2143 MEDLINE
(10) Girbal-Neuhauser, E; J Immunol 1999, V162, P585 HCAPLUS
(11) Goldbach-Mansky, R; Arthritis Res 2000, V2, P236 HCAPLUS
(12) Harding, C; J Mol Biol 1983, V70, P651
(13) Hassfeld, W; Arthritis Rheum 1995, V38, P777 MEDLINE
(14) Hunkapiller, M; Methods Enzymol 1983, V91, P227 HCAPLUS
(15) Janssens, X; J Rheumatol 1988, V15, P1346 MEDLINE
(16) Konigsberg, W; Methods Enzymol 1983, V91, P254 HCAPLUS
(17) Lichtenstein, M; J Rheumatol 1991, V18, P989 MEDLINE
(18) Lynley, A; Biochim Biophys Acta 1983, V744, P28 HCAPLUS
(19) Masi, A; Arch Intern Med 1983, V43, P2167
(20) Masson-Bessiere, C; Clin Exp Immunol 2000, V119, P544 HCAPLUS
(21) Masson-Bessiere, C; J Immunol 2001, V166, P4177 HCAPLUS
(22) Meheus, L; Clin Exp Rheumatol 1999, V17, P205 MEDLINE
(23) Munthe, E; Clin Exp Immunol 1972, V12, P55 MEDLINE
(24) Paimela, L; Ann Rheum Dis 2001, V60, P32 HCAPLUS
(25) Peterson, G; Methods Enzymol 1983, V91, P95 HCAPLUS
(26) Pincus, T; J Rheumatol 1994, V21, P1385 MEDLINE
(27) Sakata, A; Clin Exp Immunol 1996, V104, P247 MEDLINE
(28) Schellekens, G; Arthritis Rheum 2000, V43, P155 HCAPLUS
(29) Schellekens, G; J Clin Invest 1998, V101, P273 HCAPLUS
(30) Sebbag, M; J Clin Invest 1995, V95, P2672 HCAPLUS
(31) Senshu, T; J Invest Dermatol 1995, V105, P163 HCAPLUS
(32) Simon, M; J Clin Invest 1993, V92, P1387 HCAPLUS
(33) Trentham, D; J Clin Invest 1978, V62, P359 MEDLINE
(34) Utz, P; Arthritis Rheum 1998, V41, P1152 HCAPLUS
(35) Utz, P; J Exp Med 1997, V185, P843 HCAPLUS
(36) van Jaarsveld, C; Clin Exp Rheumatol 1999, V17, P689 MEDLINE
(37) Vincent, C; Ann Rheum Dis 1999, V58, P42 MEDLINE
(38) Vincent, C; J Rheumatol 1998, V25, P838 HCAPLUS
(39) Westgeest, A; J Rheumatol 1987, V14, P893 MEDLINE
(40) Wood, D; J Biol Chem 1989, V264, P5121 HCAPLUS

L28 ANSWER 2 OF 3 HCAPLUS COPYRIGHT 2002 ACS
AN 2002:1017 HCAPLUS
DN 136:384829
TI Specific presence of intracellular **citrullinated** proteins in rheumatoid arthritis synovium: Relevance to antifilaggrin autoantibodies
AU Baeten, Dominique; Peene, Isabelle; **Union, Ann; Meheus, Lydie**; Sebbag, Mireille; Serre, Guy; Veys, Eric M.; De Keyser, Filip
CS Ghent University, Ghent, Belg.
SO Arthritis & Rheumatism (2001), 44(10), 2255-2262
CODEN: ARHEAW; ISSN: 0004-3591
PB Wiley-Liss, Inc.
DT Journal
LA English
CC 15-8 (Immunochemistry)
AB To investigate the presence of **citrullinated** proteins in the synovial membrane of patients with rheumatoid arthritis (RA) and controls, and to analyze a possible relationship with antifilaggrin auto-antibody (AFA) reactivity. Synovial biopsy samples were obtained from 88 consecutive patients undergoing needle arthroscopy for knee synovitis assocd. with RA (n = 36), spondylarthropathy (n = 35), osteoarthritis (n = 9), or other diagnoses (n = 8). Tissue sections were stained with 2 different **anticitrulline** polyclonal antibodies and an antifilaggrin monoclonal antibody (mAb). The phenotype of **citrulline**-pos. cells and the colocalization with affinity-purified AFA were investigated by double immunofluorescence on frozen sections. Studies with the first antibody showed that

citrulline is expressed intracellularly in the lining and sublining layers of RA synovial tissue. Staining with the second antibody, monospecific for proteins contg. modified **citrulline**, and with anti-inducible nitric oxide synthetase confirmed the presence of **citrullinated** proteins rather than free **citrulline** in the synovium. **Citrulline**-pos. cells were detected in 50% of the RA patients (18 of 36) but in none of the controls (0 of 52). The **anticitrulline** reactivity colocalized with affinity-purified AFA reactivity, although stainings with the antifilaggrin mAb indicated the absence of filaggrin in the synovium. Intracellular **citrullinated** proteins, which are not recognized by an antifilaggrin mAb, are expressed in RA but not in control synovium. The high specificity of this finding and the colocalization with AFA reactivity boost the interest in **citrullinated** proteins as possible triggers of autoimmune responses in RA. Moreover, this is the first description of a specific histol. marker for RA synovium.

ST human rheumatoid arthritis **citrullinated** protein synovium antifilaggrin autoantibody

IT Antibodies
RL: ADV (Adverse effect, including toxicity); DGN (Diagnostic use); BIOL (Biological study); USES (Uses)
(autoantibodies; intracellular **citrullinated** proteins in rheumatoid arthritis synovium relevance to antifilaggrin autoantibodies)

IT Biomarkers (biological responses)
Human
Rheumatoid arthritis
Synovial membrane
(intracellular **citrullinated** proteins in rheumatoid arthritis synovium relevance to antifilaggrin autoantibodies)

IT Proteins
RL: ADV (Adverse effect, including toxicity); DGN (Diagnostic use); BIOL (Biological study); USES (Uses)
(intracellular **citrullinated** proteins in rheumatoid arthritis synovium relevance to antifilaggrin autoantibodies)

IT **372-75-8, Citrulline**
RL: ADV (Adverse effect, including toxicity); DGN (Diagnostic use); BIOL (Biological study); USES (Uses)
(intracellular **citrullinated** proteins in rheumatoid arthritis synovium relevance to antifilaggrin autoantibodies)

RE.CNT 31 THERE ARE 31 CITED REFERENCES AVAILABLE FOR THIS RECORD

RE

(1) Altman, R; Arthritis Rheum 1986, V29, P1039 MEDLINE
(2) Arnett, F; Arthritis Rheum 1988, V31, P315 MEDLINE
(3) Asaga, H; Biochem Biophys Res Commun 1998, V243, P641 HCAPLUS
(4) Baeten, D; Ann Rheum Dis 2000, V59, P945 MEDLINE
(5) Baeten, D; Clin Rheumatol 1999, V18, P434 MEDLINE
(6) Blass, S; Ann Rheum Dis 1998, V57, P220 HCAPLUS
(7) Brahms, H; J Biol Chem 2000, V275, P17122 HCAPLUS
(8) Despres, N; J Rheumatol 1994, V21, P1027 HCAPLUS
(9) Dougados, M; Arthritis Rheum 1991, V34, P1218 MEDLINE
(10) Girbal, E; Ann Rheum Dis 1993, V52, P749 HCAPLUS
(11) Girbal-Neuhauser, E; J Immunol 1999, V162, P585 HCAPLUS
(12) Goldbach-Mansky, R; Arthritis Res 2000, V2, P236 HCAPLUS
(13) Guerassimov, A; Arthritis Rheum 1998, V41, P1019 HCAPLUS
(14) Hoet, R; Ann Rheum Dis 1991, V50, P611 MEDLINE
(15) Janssens, X; J Rheumatol 1988, V15, P1346 MEDLINE
(16) Kraan, M; Rheumatology (Oxford) 1999, V38, P1074 MEDLINE
(17) Masson-Bessiere, C; Clin Exp Immunol 2000, V119, P544 HCAPLUS
(18) Masson-Bessiere, C; J Immunol 2001, V166, P4177 HCAPLUS
(19) Pozza, M; J Rheumatol 2000, V27, P1121 HCAPLUS
(20) Schellekens, G; J Clin Invest 1998, V101, P273 HCAPLUS
(21) Sebbag, M; J Clin Invest 1995, V95, P2672 HCAPLUS

(22) Senshu, T; Anal Biochem 1992, V203, P94 HCAPLUS
(23) Senshu, T; Biochem Biophys Res Commun 1996, V225, P712 HCAPLUS
(24) Senshu, T; J Invest Dermatol 1995, V105, P163 HCAPLUS
(25) Simon, M; Clin Exp Immunol 1995, V100, P90 MEDLINE
(26) Simon, M; J Clin Invest 1993, V92, P1387 HCAPLUS
(27) Utz, P; Arthritis Rheum 1998, V41, P1152 HCAPLUS
(28) Verheijden, G; Arthritis Rheum 1997, V40, P1115 HCAPLUS
(29) Vincent, C; Ann Rheum Dis 1999, V58, P42 MEDLINE
(30) Vincent, C; J Rheumatol 1998, V25, P838 HCAPLUS
(31) Williams, D; Rheumatology, 1st ed 1994, P9.1

L28 ANSWER 3 OF 3 HCAPLUS COPYRIGHT 2002 ACS
AN 1999:380965 HCAPLUS
DN 131:31040
TI Synthetic peptides containing **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment
IN **Meheus, Lydie; Union, Ann;** Raymackers, Joseph
PA **Innogenetics N.V., Belg.**
SO PCT Int. Appl., 74 pp.
CODEN: PIXXD2
DT Patent
LA English
IC ICM C07K014-47
ICS C07K001-107; C07K016-18; A61K038-17; G01N033-564
CC 15-2 (Immunochemistry)
Section cross-reference(s): 3

FAN.CNT 1

	PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
PI	WO 9928344	A2	19990610	WO 1998-EP7714	19981130
	WO 9928344	A3	19990812		
	W: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM				
	RW: GH, GM, KE, LS, MW, SD, SZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG				
	EP 949270	A1	19991013	EP 1998-870078	19980409
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
	CA 2309534	AA	19990610	CA 1998-2309534	19981130
	AU 9921558	A1	19990616	AU 1999-21558	19981130
	EP 1034186	A2	20000913	EP 1998-965715	19981130
	R: AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, SE, MC, PT, IE, SI, LT, LV, FI, RO				
	JP 2002512939	T2	20020508	JP 2000-523235	19981130
PRAI	EP 1997-870195	A	19971128		
	EP 1998-870078	A	19980409		
	WO 1998-EP7714	W	19981130		

AB The present invention relates to a method of producing certain peptides contg. **citrulline** residues that constitute immunogenic determinants of antibodies present in sera from patients with rheumatoid arthritis and wherein the presence of at least one **citrulline** is a prerequisite for reacting with said antibodies. The invention also relates to a method of producing said antibodies and the use of said peptides for diagnosis and treatment of rheumatoid arthritis. The **citrulline**-contg. peptides, may be circularized or branched peptides and/or contg. tandem repeats, are derived from variant of filaggrin, intermediate filament protein, vimentin, cytokeratin 1 or cytokeratin 9.

ST filaggrin intermediate filament protein vimentin cytokeratin; autoantigen

autoantibody rheumatoid arthritis autoimmune disease; antibody antiidiotype immunotoxin autoimmune disease tolerance

IT Keratins
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(1; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Keratins
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(9; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Antibodies
(anti-idiotypic; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Antibodies
RL: ADV (Adverse effect, including toxicity); BPR (Biological process); BSU (Biological study, unclassified); BIOL (Biological study); PROC (Process)
(autoantibodies; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Antigens
RL: ADV (Adverse effect, including toxicity); BPR (Biological process); BSU (Biological study, unclassified); BIOL (Biological study); PROC (Process)
(autoantigens; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Peptides, biological studies
RL: BPN (Biosynthetic preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(**citrulline**-contg.; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Toxins
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(conjugates, **citrulline**-contg. peptide; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Peptides, biological studies
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(cyclic, **citrulline**-contg.; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Test kits
(diagnostic; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Lupus erythematosus
(discoid; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Immunoassay
(enzyme-linked immunosorbent assay; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Bacteria (Eubacteria)
Eukaryote (Eukaryotae)
Yeast
(host; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Drug delivery systems
(immunotoxins; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Proteins, specific or class
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(intermediate filament-assocd.; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Antibodies
RL: BPN (Biosynthetic preparation); THU (Therapeutic use); BIOL (Biological study); PREP (Preparation); USES (Uses)
(monoclonal; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Gene
RL: BSU (Biological study, unclassified); BIOL (Biological study)
(regulatory; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Connective tissue
(scleroderma; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Membranes, nonbiological
(strip; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Autoimmune disease
Baculoviridae
Bioassay
Blood serum
Dermatomyositis
Drug screening
Immune tolerance
Immunoassay
Molecular cloning
Protein sequences
Rheumatoid arthritis
Sjogren's syndrome
Vaccines
(synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Filaggrin
Vimentins
RL: BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Immune complexes
RL: ADV (Adverse effect, including toxicity); BSU (Biological study, unclassified); REM (Removal or disposal); BIOL (Biological study); PROC (Process)
(synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera for increasing size and clearance of immune complexes in rheumatoid arthritis sera)

IT Lupus erythematosus
(systemic; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Repetitive DNA
RL: BSU (Biological study, unclassified); BIOL (Biological study)
(tandem; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT Medical goods
(test strip; synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT **372-75-8, Citrulline**
RL: BPR (Biological process); BSU (Biological study, unclassified); THU (Therapeutic use); BIOL (Biological study); PROC (Process); USES (Uses)
(synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT 75536-80-0, Peptidylarginine deiminase
RL: BSU (Biological study, unclassified); BIOL (Biological study)
(synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

IT 226904-10-5 226904-13-8 226904-18-3 226904-22-9 226904-27-4
226904-31-0 226904-37-6 226904-43-4
RL: BSU (Biological study, unclassified); PRP (Properties); THU (Therapeutic use); BIOL (Biological study); USES (Uses)
(synthetic peptides contg. **citrulline** recognized by rheumatoid arthritis sera as tools for diagnosis and treatment)

=> fil biosis
FILE 'BIOSIS' ENTERED AT 15:28:11 ON 26 AUG 2002

FILE COVERS 1969 TO DATE.
CAS REGISTRY NUMBERS AND CHEMICAL NAMES (CNs) PRESENT FROM JANUARY 1969 TO DATE.

RECORDS LAST ADDED: 21 August 2002 (20020821/ED)

=> d all tot

L38 ANSWER 1 OF 2 BIOSIS
AN 2001:537368 BIOSIS
DN PREV200100537368
TI HLA DR shared epitope, rheumatoid factor, anti-perinuclear factor, antifilaggrin and anti-cyclic **citrullinated** peptide antibodies in patients with longstanding rheumatoid arthritis: Relation with radiological progression.
AU Peene, I. (1); Kruithof, E. (1); **Union, A.**; **Meheus, L.** ; Mielants, H. (1); Veys, E. M. (1); De Keyser, F. (1)
CS (1) Dept. of Rheumatology, Ghent University Hospital, Ghent Belgium
SO Clinical Rheumatology, (2001) Vol. 20, No. 5, pp. 397. print.
Meeting Info.: 5th Belgian Congress on Rheumatology Hasselt, Belgium September 27-29, 2001
ISSN: 0770-3198.
DT Conference
LA English
SL English
CC General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals *00520
Radiation - Radiation and Isotope Techniques *06504
Clinical Biochemistry; General Methods and Applications *10006
Bones, Joints, Fasciae, Connective and Adipose Tissue - Pathology *18006
Immunology and Immunochemistry - Immunopathology, Tissue Immunology *34508
Allergy *35500
BC Hominidae 86215
IT Major Concepts
Clinical Chemistry (Allied Medical Sciences); Rheumatology (Human Medicine, Medical Sciences)
IT Diseases
rheumatoid arthritis: connective tissue disease, immune system disease, joint disease
IT Chemicals & Biochemicals

HLA DR shared epitope; anti-cyclic **citrullinated** peptide antibodies; anti-perinuclear factor; antifilaggrin antibodies; rheumatoid factor
IT Alternate Indexing
Arthritis, Rheumatoid (MeSH)
IT Methods & Equipment
radiology: analytical method
IT Miscellaneous Descriptors
joint damage progression; Meeting Abstract
ORGN Super Taxa
Hominidae: Primates, Mammalia, Vertebrata, Chordata, Animalia
ORGN Organism Name
human (Hominidae): patient
ORGN Organism Superterms
Animals; Chordates; Humans; Mammals; Primates; Vertebrates

L38 ANSWER 2 OF 2 BIOSIS COPYRIGHT 2002 BIOLOGICAL ABSTRACTS INC.
AN 1998:468693 BIOSIS
DN PREV199800468693
TI Epitope mapping of natural filaggrin leads to the identification of rheumatoid arthritis-immunoreactive epitopes containing **citrulline**.
AU **Union, Ann (1)**; Amerijckx, Liesbet (1); Raymackers, Jos (1); Dauwe, Martine (1); De Keyser, Filip; Veys, Eric; **Meheus, Lydie (1)**
CS (1) **Innogenetics** N.V., Industriepark 7, 9052 Ghent Belgium
SO Arthritis & Rheumatism, (Sept., 1998) Vol. 41, No. 9 SUPPL., pp. S84. Meeting Info.: 62nd National Scientific Meeting of the American College of Rheumatology and the 33rd National Scientific Meeting of the Association of Rheumatology Health Professionals San Diego, California, USA November 8-12, 1998 American College of Rheumatology . ISSN: 0004-3591.
DT Conference
LA English
CC Biochemical Studies - General *10060
Immunology and Immunochemistry - General; Methods *34502
General Biology - Symposia, Transactions and Proceedings of Conferences, Congresses, Review Annuals *00520
IT Major Concepts
Biochemistry and Molecular Biophysics
IT Diseases
rheumatoid arthritis: connective tissue disease, immune system disease, joint disease
IT Chemicals & Biochemicals
citrulline; filaggrin; rheumatoid arthritis-immunoreactive epitopes
IT Methods & Equipment
epitope mapping: analytical method
IT Miscellaneous Descriptors
Meeting Abstract; Meeting Poster
RN **372-75-8 (CITRULLINE)**

=> d his

(FILE 'HOME' ENTERED AT 15:05:52 ON 26 AUG 2002)
SET COST OFF

FILE 'REGISTRY' ENTERED AT 15:06:42 ON 26 AUG 2002
E QDTIHGHPCS'CIT''CIT'GCRPGY/SQEP
L1 1 S E1
L2 1 S QDTIHGHPCS..GCRPGY/SQSP

```
L3            1 S L1,L2
L4            0 S [STDENQHKR][STDENQHKR][STDENQHKR][ILVAM][HKRDESTYFW]G[HKRDEST
L5            1 S [STDENQHKR][STDENQHKR][STDENQHKR][ILVAM][HKRDESTYFW]G[HKRDEST
L6            1 S L3,L5
                SAV L6 DIBRINO747/A

     FILE 'HCAOLD' ENTERED AT 15:12:44 ON 26 AUG 2002
L7            0 S L6

     FILE 'USPATFULL, USPAT2' ENTERED AT 15:12:47 ON 26 AUG 2002
L8            0 S L6

     FILE 'HCAPLUS' ENTERED AT 15:12:52 ON 26 AUG 2002
L9            1 S L6
                SEL RN

     FILE 'REGISTRY' ENTERED AT 15:13:12 ON 26 AUG 2002
L10          19 S E1-E19
L11          18 S L10 NOT L6
L12           3 S L11 NOT SQL/FA
L13          15 S L11 NOT L12
L14          10 S L13 AND SQL<=18

     FILE 'HCAPLUS' ENTERED AT 15:17:39 ON 26 AUG 2002
                E UNION A/AU
L15          11 S E3,E4
                E MOEREELS H/AU
L16          55 S E3,E4,E6-E8
                E MEHEUS L/AU
L17          25 S E3-E5
                E INNOGENET/PA,CS
L18         166 S E4-E40
L19           1 S L15-L18 AND L9
L20         240 S L15-L18 NOT L19
L21           3 S L20 AND ?CITRUL?

     FILE 'REGISTRY' ENTERED AT 15:20:39 ON 26 AUG 2002
L22           1 S L12 AND ORNITH?
                E D-CITRULLINE/CN
L23           1 S E3
                E DL-CITRULLINE/CN
L24           1 S E3

     FILE 'HCAPLUS' ENTERED AT 15:21:06 ON 26 AUG 2002
L25           3 S L22 AND L15-L18
L26           4 S L21,L25
L27           4 S L26 AND ?CITRUL?
L28           3 S L27 NOT L19
                SEL RN

     FILE 'REGISTRY' ENTERED AT 15:21:46 ON 26 AUG 2002
L29          10 S E1-E10
L30           1 S L29 AND N5
L31           1 S L29 AND AMINOCARBONY?
L32           1 S L29 AND ORNITH?
L33           1 S L30-L32

     FILE 'REGISTRY' ENTERED AT 15:24:23 ON 26 AUG 2002

     FILE 'HCAPLUS' ENTERED AT 15:25:02 ON 26 AUG 2002

     FILE 'BIOSIS' ENTERED AT 15:26:03 ON 26 AUG 2002
                E UNION A/AU
```

```
L34          13 S E3,E4
                E MOEREELS H/AU
L35          40 S E3-E7
                E MEHEUS L/AU
L36          26 S E3-E6
                E INNOGENET/CS
                E INNOGEN/CS
L37         208 S E3-E85
L38           2 S L34-L37 AND (L6,L22-L24 OR ?CITRUL?)

     FILE 'BIOSIS' ENTERED AT 15:28:11 ON 26 AUG 2002
```